COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA

SOBRE OS AUTORES

George Barreto **Bezerra** é pesquisador do Laboratório de Bio-Informática e Computação Bio-Inspirada (LBiC), Faculdade de Engenharia Elétrica e de Computação (DCA-FEEC-UNICAMP).

Leandro N. de **Castro** <lnunes@unisantos.br> é professor do Programa de Mestrado em Informática, da Universidade Católica de Santos (UniSantos).

Andre **De Tienne** <adetienn@iupui.edu> é professor do Departamento de Filosofia da IUPUI, Indianápolis, e editor associado ao Peirce Edition Project.

Charbel **El-Hani** <charbel@ufba.br> é professor e coordenador do Programa de Pós-Graduação em Ensino, Filosofia e História das Ciências, Instituto de Biologia (UFBA); e do Programa de Pós-Graduação em Ecologia e Biomonitoramento (UFBA).

Claus **Emmeche** <emmeche@nbi.dk> é professor e diretor do Centro de Filosofia da Naureza da Universidade de Copenhagen.

Ricardo R. **Gudwin** <gudwin@dca.fee.unicamp.br> é professor do Departamento de Engenharia de Computação e Automação Industrial da Faculdade de Engenharia Elétrica e de Computação (DCA-FEEC- UNICAMP).

W. (Pim) F.G. **Haselager** <pimh@nici.kun.nl> é professor no Instituto de Cognição e Informação da Universidade de Nijmegen, e do Departamento de Filosofia da Universidade Estadual de São Paulo (UNESP, Marília).

Angelo **Loula** <angelocl@dca.fee.unicamp.br> é professor da Área de Informática no Departamento de Ciências Exatas da Universidade Estadual de Feira de Santana (UEFS).

Alexander **Mehler** <mehler@uni-trier.de> é professor e assistente científico de linguística computacional e processamento de dados linguísticos da Universidade de Trier.

Winfried **Nöth** <wnoth@pucsp.br> é professor na Univesidade de Kassel, diretor do Centro Interdisciplinar para Estudos Culturais, na mesma universidade.

João **Queiroz** <queirozj@semiotics.pro.br> <queirozj@gmail.com> é professor e pesquisador do Programa de Pós-Graduação em Ensino, Filosofia e História das Ciências (UFBA), Instituto de Biologia (UFBA), e do Programa de Pós-Graduação em Ecologia e Biomonitoramento (UFBA).

Joseph **Ransdell** <joseph.ransdell@yahoo.com> é professor emérito da Texas Tech University.

Lucia **Santaella** Braga <lbraga@pucsp.br> é professora e coordenadora do Programa de Pós-graduação em Tecnologias da Inteligência e Design Digital (PUC/SP).

Janaína Stella de **Sousa** é pesquisadora do Laboratório de Bio-Informática e Computação Bio-Inspirada (LBiC), Faculdade de Engenharia Elétrica e de Computação (DCA-FEEC-UNICAMP).

Tom **Ziemke** <tom@ida.his.se> é professor de Ciência Cognitiva no departamento de Ciência de Computação da Universidade de Skovde

SUMÁRIO

Sobre os autores **07**

Introdução **09**
Sobre a síntese de sistemas e criaturas semióticas
João Queiroz

Capítulo 1 **19**
A relevância da semiótica Peirceana para uma inteligência computacional aumentada
Joseph Ransdell

Capítulo 2 **67**
Aprendizagem *qua* semiose
André De Tienne

Capítulo 3 **93**
Estruturalismo hierárquico, semiose e emergência
Charbel Niño El-Hani e João Queiroz

Capítulo 4 **129**
O que é o símbolo
Lucia Santaella

Capítulo 5 **145**
Aspectos metodológicos da semiótica computacional
Alexander Mehler

Capítulo 6 **159**
Máquinas semióticas
Winfried Nöth

Capítulo 7 **185**
Um robô possui *Umwelt*? Reflexões sobre a Biossemiótica Qualitativa de Jakob von Uexküll
Claus Emmeche

Capítulo 8 **235**
Robossemiótica, cognição enativa e incorporada
Tom Ziemke

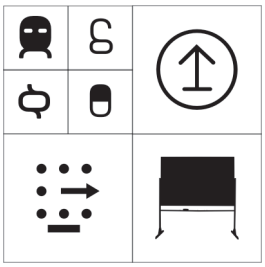
Capítulo 9 **249**
Forma, função e a matéria da experiência
Pim Haselager

Capítulo 10 **265**
Engenharia imunológica e cognição: da natureza à solução de problemas de engenharia
Leandro Nunes de Castro, Janaína Stella de Sousa, George Barreto Bezerra

COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA



INTRODUÇÃO

SOBRE A SÍNTESE DE SISTEMAS E CRIATURAS SEMIÓTICAS

João Queiroz

O slogan ‘construir para explicar’ assume, com os computadores digitais, um sentido inédito na história das ciências, e é hoje considerado uma coluna vertebral de disciplinas e departamentos. Sistemas e criaturas computacionais de todo tipo são implementados em diferentes plataformas, por meio de muitas técnicas, e motivados por diversos objetivos. Em contra-partida, para a teoria simulada, já que toda simulação traduz uma teoria para linguagem de programação (Parisi 2001), significa uma oportunidade de quantificar e formalizar suas asserções. Além disso, simulações fornecem meios inéditos para realização de ‘experimentos mentais’ dos fenômenos investigados (Bedau 1998, Dennett 1998): como seriam, ou teriam sido, tais e tais fenômenos, se as condições para a emergência e desenvolvimento fossem, ou tivessem sido, outras, e não estas?

Em termos experimentais, são muitas as vantagens defendidas: pode-se alterar livremente os parâmetros que definem os padrões de eventos observados, a arquitetura dos sistemas, o ambiente e

as leis que regem o comportamento dos objetos simulados; pode-se isolar e variar cada parâmetro isoladamente, associar diversas variações, combiná-las 'em cascata' e observar as consequências decorrentes de um, ou diversos, destes procedimentos; pode-se replicar, sem as dificuldades típicas de protocolos empíricos, os procedimentos, introduzir novos e subtrair antigos parâmetros; pode-se rever a história de interação de cada sistema, ou criatura, com co-específicos, com competidores, com o ambiente e seus diversos eventos.

Quando *processos semióticos* estão em foco, as abordagens atuam em diversos níveis — sintático, morfológico, semântico, pragmático, comunicação entre criaturas, etc (Cangelosi & Parisi 2002, Steels 2003). Uma parte das abordagens simula a emergência de competências semióticas na ausência de qualquer adaptação prévia. Os sistemas são capazes de produzir alguma forma de semiose em um ambiente em que esta, seus componentes ou estruturas, não foram disponibilizados. Dependendo do quadro teórico, e das ferramentas computacionais, pode-se testar diversos fatores que afetam a ontogênese de muitos processos, como as diferenças entre sistemas de signos inatos e adquiridos, o papel adaptativo de estruturas semióticas composicionais, as vantagens decorrentes do aparecimento de processos simbólicos, os supostos substratos materiais responsáveis por estes processos, a influência entre diferentes competências semióticas (e.g. processamento simbólico) e tarefas de baixo nível cognitivo (e.g. atenção). Enfim, pode-se (e é o que se faz) experimentar 'livremente', se movendo em horizontes formais e teóricos mais ou menos consolidados, assumindo-os abertamente como meta-princípios, ou aceitando-os tacitamente como 'fontes de inspiração'.

Vida artificial, robótica cognitiva, ANIMATS, etologia sintética e semiótica computacional estão entre as principais áreas envolvidas na construção de sistemas e criaturas semióticas artificiais. Elas se baseiam no uso de diferentes ferramentas, e divergem em muitas de suas pretenções, mas são fortemente influenciadas por meta-princípios (*formal-theoretical constraints*) e por motivações empíricas (*empirical constraints*), para o design dos ambientes e definição dos sistemas, como morfologia de sensores, efetores, arquitetura e processos cognitivos das criaturas concebidas. Na prática, isto significa que dois conjuntos de restrições informam ao cientista: o que ele pretende simular? o que precisa ser considera-

do? como saber (critérios de avaliação) se o resultado é uma boa simulação?

As relações entre teorias, modelos e simulações são 'vias de mão dupla'. *Restrições teóricas*, como aquelas derivadas dos modelos de Jakob von Uexkull, e da semiótica e pragmatismo de C.S.Peirce, combinadas a descrições de fenômenos físicos e biológicos, têm fornecido subsídios, provocações, além de uma bateria de fenômenos para modelar e simular. Boa parte do *background* teórico encontrado aqui se baseia nas obras de C.S. Peirce e de Jakob von Uexkull. Peirce é considerado, com Frege, Russell, e Hilbert, um dos fundadores da lógica moderna (Lukasiewicz 1970: 111; Barwise & Etchemendy 1995: 211; Quine 1995: 23; Hintikka & Hilpinen 1997: ix). Uexkull é um dos fundadores da etologia (ver Kull 2001).

Peirce também é considerado o fundador da moderna teoria do signo, ou semiótica. Ele a desenvolve em um ambiente bastante formal de especulação, baseado em uma teoria lógica-fenomenológica de categorias. A semiótica é definida por Peirce como a 'doutrina da natureza essencial e fundamental de todas as variedades de possíveis semioses' (CP 5.484).[1] Os conceitos peirceanos mais recorrentes que o leitor encontrará neste livro são os de signo, semiose, e suas variações em muitas classes (ícone, índice e símbolo; qualisigno, sinsigno, legisigno, etc). A semiose (ou 'ação do signo') é descrita como uma relação triádica irredutível entre signo, objeto e interpretante (efeito do signo). Este modelo tem influenciado muitos autores, e diversas comunidades científicas (ver Vogt 2002, 2007; Pietarinen 2005; Freadman 2004; Queiroz & Merrell 2005; Deacon 1997; Fetzer 1997; Houser 1997; Hoffmeyer 1996; Habermas 1995; Noble & Davidson 1996; Emmeche 1991; Fisch 1986). Trata-se de um modelo relacional, dinâmico, contexto e intérprete-dependente. A irredutibilidade lógica da tríade (signo-objeto-interpretante, S-O-I), sua indecomponibilidade triádica, indica que a relação depende constitutivamente dos três termos.

Peirce define, pragmaticamente, o signo como um 'meio para a comunicação de uma forma', ou um hábito, incorporado no objeto, de tal modo a restringir o comportamento de um intérprete (Bergman 2000 a,b). É uma questão empírica, muito dependente de pressupostos fundamentais, *se*, e *quais*, criaturas (ou sistemas) biológicos ou artificiais, são capazes de usar signos para 'comunicar formas' (padrões de similaridade, ou ícones, correlações espaço-temporais, ou índices, relações legaliformes, ou símbolos) in-

corporadas em objetos, de modo a constrangir o comportamento de intérpretes. A questão tanto é um desafio às discussões sobre os limites da 'semiose genuina', em sistemas artificiais e em criaturas biológicas não-humanas, quanto sobre a existência de *Umwelt* nestes sistemas e criaturas.

O leitor encontrará diversas alusões às intrincadas classificações dos signos de Peirce (especialmente o capítulo de Santaella, mas também de Emmeche, Haselager, Nöth). Elas têm atraído gerações de filósofos, linguistas, semioticistas e, mais recentemente, de biólogos, roboticistas e lógicos. O modelo triádico de semiose não especifica a natureza (categorial) dos termos S-O-I, e não especifica a natureza (categorial) das relações entre S, O e I. As classificações sígnicas respondem às perguntas: (I) quantas 'variedades fundamentais' (CP 5.488) podem ser concebidas? (ii) quais são estas variedades? (iii) como elas estão relacionadas? Relativamente à 'mais fundamental divisão de signos', as categorias aproximadamente correspondem a ícones, índices e símbolos. Esta classificação é bem conhecida, tem sido utilizada por muitos autores, em diversas áreas, e descreve as relações que se podem estabelecer entre os signos e seus objetos. Pressionado por descobertas em diferentes domínios (teoria dos grafos, fenomenologia), Peirce desenvolveu diversas classificações sígnicas. Elas permitem uma descrição bastante detalhada das relações que operam na tríade S-O-I. Como exemplo, as dez classes de signos, desenvolvidas a partir de 1903, permitem responder às questões: (I) qual a natureza do signo? (ii) qual a natureza da relação signo-objeto? (iii) qual a natureza da relação entre o signo e seu objeto para seu interpretante? Um signo pode ser uma qualidade (qualisigno), uma ocorrência (sinsigno), ou uma lei (legisigno); pode estar relacionado por similaridade com seu objeto (ícone), por correlação espaço-temporal (índice), ou através de uma convenção ou hábito (símbolo); pode ser interpretado como uma hipótese (rema), como um designador (dicente) ou como uma regra (argumento).

As classificações sígnicas não representam apenas 'reduções' de variados eventos semióticos a complicadas tipologias. Elas conectam uma variedade concebível de eventos por meio de princípios gerais estabelecidos em um ambiente lógico-fenomenológico de descrição e análise. Uma vez que a preocupação primária de Peirce, como lógico e matemático, não fora com o signo linguístico, suas descrições não sofreram do linguicentrismo típico de conhecidas verten-

tes semiológicas, a partir das quais pouco pode-se fazer, ao examinar um robô ou uma criatura, que subtrair-lhes propriedades semióticas complexas (e.g. sintaxe, composicionalidade semântica etc), e, uma vez que as discussões sobre *Umwelt* e semiose parecem estar indissociavelmente conectadas, que subtrair-lhes 'mundo fenomenal', ou *Umwelt*.

Para J.Uexkull, o que é cognitivamente significativo para uma criatura depende de sua interação sensório-motora com a informação disponível em seu ambiente. *Umwelt* pode ser definido como o aspecto fenomenal das partes do ambiente de uma espécie. As partes que a espécie, evolutivamente, 'escolhe' em termos sensório-motores, de acordo com sua organização e suas necessidades.

É crescente o número de trabalhos sobre *Umwelt* em etologia, biossemiótica, filosofia da biologia, além de vida artificial, e pesquisas sobre sistemas autônomos. A questão, retomada aqui em diversas ocasiões, é: uma criatura artificial pode viver em um mundo fenomenal, de acordo com a noção de *Umwelt*? É exatamente esta a questão à que Claus Emmeche dedica sua atenção: robôs têm, ou podem ter, *Umwelt*? Emmeche defende a noção de *Umwelt* como particularmente relevante para a *nouvelle* IA, uma vez que ela enfatiza a interação que decorre da experiência. O capítulo de Pim Haselager é também uma discussão sobre o papel do *Umwelt* em criaturas artificias. Ele questiona a relação de co-dependência, estabelecida por diversos pesquisadores, entre *Umwelt* e vida, no contexto da robótica cognitiva situada e incorporada. Tom Ziemke discute a possibilidade de efetiva implementação de semiose artificial em agentes autônomos. Ele apresenta 'agentes autônomos' como modelos de processos sígnicos, e cognição enativa incorporada. Ziemke está interessado nas propriedades de autonomia e capacidade de semiose destes agentes.

Uma discussão detalhada sobre sistemas autônomos situados, ou sobre sistemas semióticos auto-organizados com propriedades qualitativas emergentes, ainda está por ser feita. A noção de emergência raramente é discutida em IA e em vida artificial. O capítulo de El-Hani & Queiroz discute em que sentido a semiose pode ser caracterizada como um processo 'emergente'. O problema está relacionado às condições que precisam ser satisfeitas para tal caracterização. Os autores propõem um modelo capaz de explicar 'emergência de semiose' em sistemas que produzem, processam e

interpretam signos, baseado no estruturalismo hierárquico de Stanley Salthe.

Winfried Nöth aborda a noção de 'máquina semiótica', sua relação com as noções de semiose, e, especialmente, de quasi-semiose, definida como uma forma de semiose não-genuína. (Formas genuínas de semiose não devem se basear em procedimentos mecânicos ou em relações causais de eficiência.) Nöth sugere, com a tese sinequista de Peirce de pano-de-fundo, que há um gradiente semiótico em máquinas de diversos tipos. Esta posição lhe permite descrever processos mecânicos, quasi-mentais, cujos atributos podem ser identificados em mentes (e.g. quando o raciocínio opera mecanicamente) e em máquinas (quando elas exibem auto-controle). A associação entre semiose genuina e processos *autopoieticos* (máquinas auto-organizadas), confere ao tratamento de Nöth um lugar de destaque nas dicussões sobre o *fundamento do símbolo*, e *autonomia*, em inteligência computacional e vida artificial.

Há duas áreas em inteligência computacional que devem ser dintinguidas em seus objetivos e pretensões — inteligência artificial (IA) e inteligência aumentada. Elas são complementares. Joseph Ransdell está interessado na exploração da segunda, cujo propósito é regular ou coordenar aspectos mecanizáveis da inteligência, expandindo-a. A área não está interessada em um modelo da inteligência, uma 'vertente' que tem na Máquina de Turing, e no Teste de Turing, seus principais modelos. Vannevar Bush, é o autor-chave aqui, e MEMEX é a máquina-modelo. Ransdell desenvolve a noção de Skagestad sobre inteligência aumentada, com ênfase nos aspectos dialógicos de processos sígnicos materialmente incorporados. Seu texto explora as noções de inteligência computacional, de mente como 'prática comunicacional' e discute o papel da semiótica de Peirce como *framework* para tratá-los. Para Skagestad a semiótica de Peirce fornece as bases conceituais mais adequadas para entender e consolidar uma tradição de pesquisas em inteligência aumentada. Ransdell, que está de acordo com essa visão, analisa um caso (Sistema Ginsparg) em que técnicas computacionais são usadas para implementar um controle crítico de publicações científicas, com foco em processos de agenciamento das práticas envolvidas na atividade científica de publicação.

Andre DeTienne examina a noção de aprendizagem como um processo temporal, regulado por princípios que caracterizam uma das categorias de Peirce, a Terceiridade (*Thirdness*). Ele explora as diver-

sas restrições que tal aproximação precisa satisfazer para ser bem sucedida, e estabelece cinco princípios de acordo com os quais Peirce descreve a aprendizagem como um fenômeno pré-psicológico.

Lúcia Santaella aborda o conceito de símbolo, para Peirce, as diversas variações e sub-divisões deste conceito, e relações com a noção de hábito. Trata-se de um tópico recorrente em ciências cognitivas, que identifica o *símbolo* com propriedades de composicionalidade e arbitrariedade semânticas, frequentemente em um sistema declarativo de sinais, propriedades às quais Peirce jamais restringiu este conceito.

O capítulo de Alexander Mehler é fortemente metodológico. Seu foco é o que se conhece hoje como semiótica computacional (ver Gudwin & Queiroz 2007). Mehler define seu escopo e sua relação com a semiótica de computadores, com a vida artificial *forte* e propõe importantes distinções entre modelagem, simulação e emulação.

Leandro de Castro e colaboradores apresentam seus desenvolvimentos em sistemas imunológicos artificiais, área em que Castro é considerado um dos principais fundadores. Estes desenvolvimentos, eles defendem, têm importantes consequências para as noções de representação, reconhecimento de padrão e informação, abrindo uma 'nova frente' nas pesquisas em vida artificial e semiótica computacional.

São abordados aqui problemas teóricos, metodológicos, e são apresentados novos modelos computacionais. Cientistas cognitivos, atuando em novos *frameworks* (e.g. *nouvelle AI*), tomam seriamente em consideração a semiótica de Peirce, e as abordagens de Uexkull. Alguns dos trabalhos deste livro discutem e desafiam a idéia de 'semiose genuína', e de *Umwelt*, em sistemas artificiais; outros, a idéia de emergência de semiose e de 'mundo fenomenal' nestes sistemas; há capítulos que definem e relacionam com precisão as noções de semiose, aprendizagem, símbolo, e máquina semiótica; e um capítulo que, inspirado no sistema imunológico, propõe novas estratégias para construção de sistemas computacionais.

AGRADECIMENTOS

Os organizadores agradecem, pela colaboração na tradução dos capítulos, a Julia Itani (Capítulo 2), Luciane Rodrigues (Capítulo 3), Antonio Gomes (Capítulo 6), Jackeline S. de Freitas (Capítulos 2 e 3), e a Virginia Dazzani pela revisão do capítulo 3. J.Q. é financiado por uma bolsa de pós-doutorado DCR (CNPq/FAPESB). A.L. agradece o apoio da FAPESB. R.G. agracede ao CNPq.

Notas

[1] A obra de Peirce será citada, neste livro, como: CP (seguido pelo número do volume e parágrafo), *The Collected Papers of Charles S. Peirce* (1866-1913); EP (seguido pelo número do volume e página), *The Essential Peirce* (1893-1913); W (seguido pelo número do volume e página), *Writings of Charles S. Peirce* (1839-1914); MS (seguido pelo número do manuscrito), *Annotated Catalogue of the Papers Of Charles S. Peirce*.

Referências

BARWISE, Jon e ETCHEMENDY, John. 1995. Heterogeneous Logic. Em: *Diagrammatic Reasoning — cognitive and computational perspective*. J. Glasgow (et al. eds.). The AAAI Press.

BERGMAN, Mats. 2000a. Reflections on the role of the communicative sign in semeiotic. *Transactions of the Charles S. Peirce Society:* A Quarterly Journal in American Philosophy, Spring XXXVI (2): 225-254.

__. 2000b. *Meaning and Mediation:* Toward a communicative interpretation of Peirce's theory of signs. Yliopistopaiano.

BRAITENBERG, Valentino. 1984. *Vehicles:* Experiments in synthetic psychology. MIT Press.

BEDAU, Mark A. 1998. Philosophical content and method of artificial life. Em: *The Digital Phoenix:* How Computers are Changing Philosophy. T.W. Bynum e J.H. Moor (eds.) Blackwell Publishers. pp.135-152.

CANGELOSI, Angelo; PARISI, Domenico (eds.). 2002. *Simulating the Evolutions of Language*. Springer Verlag.

DEACON, Terrence. (1997). *Symbolic Species:* The Co-evolution of Language and the Brain. Norton.

DENNETT, Daniel. 1998. *Brainchildren - essays on the designing minds*. MIT Press.

EMMECHE, Claus. 1991. A semiotical reflection on biology, living signs and artificial life. *Biology & Philosophy* 6 (3): 325-340.

FETZER, James. 1997. Thinking and computing: computers as special kinds of signs. *Minds and Machines* 7: 345–364.

FISCH, Max. 1986. *Peirce, semeiotic, and pragmatism*. Indiana University Press.

FREADMAN, Anne. 2004. *The Machinery of Talk – Charles Peirce and the Sign Hypothesis*. Stanford Univerity Press.

GUDWIN, Ricardo; QUEIROZ, João (eds.) 2007. *Semiotics and Intelligent Systems Development*. Idea Group Inc.

HABERMAS, J. 1995. Peirce and communication. Em: *Peirce And Contemporary Thought:* Philosophical Inquiries. K.Ketner (ed). Fordham University Press. pp. 243-266.

HINTIKKA, Jaakko; HILPINEN, Risto. 1997. Preface. Em: *Studies in the Logic of Charles S. Peirce*. N. Houser, D. Roberts, & J. Evra (eds.). Indiana University Press. pp. ix-x.

HOFFMEYER, Jesper. 1996. *Signs of Meaning in the Universe*. Indiana University Press.

HOUSER, Nathan. 1997. Introduction: Peirce as a logician. Em: *Studies in the Logic of Charles S. Peirce*. N. Houser, D. Roberts, & J. Evra (eds.). Indiana University Press. pp. 1-22.

KULL, Kalevi. (guest ed.) 2001. Jakob von Uexkull: a paradigm for biology and semiotics. *Semiotica* 134 (1/4).

LUKASIEWICZ, Jan. 1961 [1970]. *Selected Works*. (ed.) L. Borkowski. North-Holland: Amsterdan.

NOBLE, W.; DAVIDSON, I. 1996. *Human Evolution, Language and Mind*. Cambridge University Press.

STEELS, Luc. 2003. Evolving grounded communication for robots. *Trends in Cognitive Science* 7 (7): 277-324.

PARISI, Domenico. 2001. *Simulazioni – la realtà rifatta nel computer*. Il Mulino.

PIETARINEN, Ahti-Veikko. 2005. *Signs of Logic:* Peircean Themes on the Philosophy of Language, Games, and Communication (Synthese Library, Vol. 329). Springer.

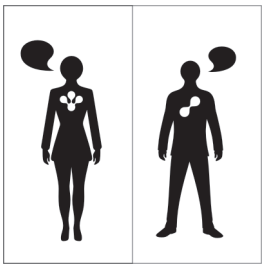
QUEIROZ, João; MERRELL, Floyd. 2005 Abduction – between subjectivity and objectivity. *Semiotica* 153 (1/4).

QUINE, Willard V. 1995. Peirce's Logic. Em: Proceeddings of the Charles S. Peirce Bicentennial International Congress. K. Ketner (et al, eds). *Texas Tech Press*. pp.23-31.

VOGT, Paul. 2002. The physical symbol grounding problem. *Cognitive Systems Research* 3(3), 429–457.

__. 2003. Anchoring of semiotic symbols. *Robotics and Autonomous Systems* 43 (2): 109-120.

__. 2007. Language evolution and robotics: issues on symbol grounding and language acquisition. Em: *Artificial Cognition Systems*. A. Loula, R. Gudwin e J. Queiroz (eds.). Idea Group Inc. pp.176-209.



CAPÍTULO 1

A relevância da semiótica Peirceana para uma inteligência computacional aumentada

Joseph Ransdell

Introdução

Peter Skagestad identifica duas visões distintas que têm estimulado as pesquisas sobre inteligência baseada em computação. Ele as chama de 'Inteligência Artificial' e 'Inteligência Aumentada' (Skagestad 1996)[1]. O objetivo deste capítulo é, em primeiro lugar, fazer a distinção entre estes dois tipos de pesquisa, em inteligência computacional, para aqueles que podem não estar acostumados a reconhecê-los como partes coordenadas. Em seguida, vou chamar a atenção para um tipo especial de pesquisa em Inteligência Aumentada, onde me parece necessária uma ênfase especial, tanto em razão de seu importante potencial quanto porque as considerações de Skagestad sobre as características distintivas da pesquisa em Inteligência Aumentada não me parecem capturar as características mais salientes deste domínio, talvez porque pode não lhe ter ocorrido que ele é suficientemente distinto para exigir atenção especial.

A pesquisa em Inteligência Artificial pode ser caracterizada como programação de computadores com o intuito de criar máquinas que possam pensar da mesma maneira, ou melhor, do que seres humanos. A pesquisa em Inteligência Aumentada, por sua vez, é a programação de computadores com o intuito de promover uma base computacional para o aumento ou incremento do pensamento humano, assistindo-o, não tentando substituí-lo por simulação em máquinas. As duas podem ser vistas como sendo, de maneira geral, complementares em suas aplicações, e o termo 'pesquisa em inteligência computacional', ou 'Pesquisa em IC', engloba ambas as áreas. O tipo particular de Inteligência Aumentada para o qual desejo chamar a atenção é a programação de computadores que se presta a apoiar, expandir e aperfeiçoar o controle de publicações e comunicações de pesquisas baseado em críticas.

Embora o trabalho de C.S.Peirce seja tão relevante para a Inteligência Artificial quanto para a Inteligência Aumentada[2], Skagestad está especialmente preocupado em situar Peirce, no que se refere ao embasamento teórico para a Inteligência Aumentada, de maneira comparável à posição fundamental de Alan Turing em relação à Inteligência Artificial, em virtude da concepção, deste último, da Máquina Universal e do famoso 'Teste de Turing' para avaliar a inteligência de computadores.

Skagestad situa Peirce desta forma explicando o que está implícito na afirmação de Peirce de que “todo pensamento é em signos”, que interpreta: todo pensamento é materialmente corporificado. Desenvolvendo a concepção de Inteligência Aumentada de Skagestad mais profundamente, na direção indicada, eu também faço uso da afirmação de Peirce, mas agora explicitando uma outra implicação (porém complementar): todo pensamento é dialógico[3]. Como um exemplo de Inteligência Aumentada desta natureza (porém não prototípico), eu utilizo o sistema servidor e os arquivos automatizados de publicações primárias criado há mais de 10 anos pelo físico Paul Ginsparg, do Laboratório Nacional de Los Alamos, e que se encontra ainda hoje em uso, com sucesso, na área de física teórica de altas energias e outras áreas na física, astronomia e matemática.

A distinção entre as pesquisas em Inteligência Artificial e em Inteligência Aumentada

Talvez os leitores deste trabalho não necessitem de referências para a literatura em pesquisas sobre a Inteligência Artificial, mas

as bases para o movimento da Inteligência Aumentada em pesquisas de inteligência computacional podem não ser igualmente familiares. A distinção está implícita na literatura especulativa sobre inteligência computacional desenvolvida nas últimas décadas. Mas o reconhecimento destes movimentos, como dois desenvolvimentos igualmente importantes, na categoria mais ampla de programação em inteligência computacional, parece ser relativamente recente[4]. Como embasamento para este artigo, recomendo três trabalhos de Peter Skagestad, sobre este tópico, que estão disponíveis on-line[5].

Os três artigos são relevantes, mas estarei aqui comentando somente alguns poucos pontos que eles levantam, principalmente (mas não exclusivamente) do artigo de 1993. Nestes artigos, Skagestad distingue entre Inteligência Artificial e Inteligência Aumentada, como tipos de metas da programação que correspondem ao que ele considera duas 'revoluções na computação', diferentes e baseadas em 'duas máquinas abstratas' bem distintas — a máquina Universal de Alan Turing, como descrita em seu artigo de 1936 sobre números computáveis, e o Memex de Vannevar Bush, como descrito no artigo de 1945. Skagestad diz:

> Ambos, a Máquina de Turing e o Memex, tentam mecanizar funções específicas da mente humana. O que Turing tentou mecanizar foi a computação e, de maneira geral, qualquer processo de raciocínio que pudesse ser representado por um algoritmo; o que Bush tentou mecanizar foram os processos associativos por meio dos quais trabalha a memória humana. [...] O Memex, que tenta replicar a memória humana, e portanto pode ser visto como corporificação de uma 'memória artificial', não tinha como intenção rivalizar com a mente humana [como faz a Inteligência Artificial] mas sim estender seu alcance, disponibilizando mais rapidamente seu conteúdo, e selecionando os registros mais úteis para uma dada situação, quando necessário. Esta idéia inspirou diretamente o programa de pesquisas conhecido como 'inteligência aumentada' (Inteligência Aumentada), formulado em 1962 por Douglas Engelbart, com um agradecimento explícito a Bush.

Skagestad observa mais adiante que:

> A máquina de Turing é o ancestral da máquina de inferência, dentro do escopo de um computador pessoal [...], enquanto o Memex de Bush é o ancestral de muitas das características a que nos referimos, coletivamente, como a interface com o usuário.

E ele nos lembra que:

> Nos anos 60, os computadores eram enormes, caros e manipuláveis somente por uma elite. A idéia de tornar estas máquinas ferramentas pessoais de gerenciamento de informação, podendo ser adquiridas de maneira generalizada e utilizadas sem treinamento especial foi defendida somente por uma minoria de visionários e era vista como bizarra não somente pelo público em geral, mas também pelas principais indústrias eletrônicas. A segunda revolução nos computadores obviamente não poderia ter acontecido sem que a primeira a precedesse, mas a primeira revolução podia facilmente ter acontecido sem ter sido seguida pela segunda.

Fenômenos de tal complexidade são freqüentemente explicáveis, de acordo com suas origens, por mais de uma perspectiva. As coisas reais têm facetas, e múltiplas perspectivas complementares, em realidades históricas complexas, são normalmente exigidas de forma a se produzir uma idéia razoavelmente sofisticada destas coisas, como um todo. Neste caso, o papel de visionários, como Turing e Bush, é sem dúvida importante, mas há outras coisas a serem ditas sobre a origem da concepção (ou concepções) do computador. Minha suposição é que, considerando a origem de sua concepção como um instrumento de uso pessoal, utilizado para aumentar as habilidades de se produzir texto, de se trabalhar com documentos de diversas maneiras, e de se comunicar com outras pessoas, esta concepção originou-se, em parte pelo menos, como um subproduto não intencional do trabalho destinado a satisfazer a necessidade de se documentar a programação usada na computação em *mainframes*, cuja manutenção exigia que registros fossem armazenados tanto para o uso próprio de um programador como também para o uso de outros programadores. Isto, por sua vez, exigia a habilidade não somente de se capturar a informação mas também de comunicá-la, o que poderia ser facilitado fazendo uso dos poderes do próprio computador, como um instrumento capaz de fazer tais registros e executar sua transmissão.

Não era, de maneira alguma, necessário que se fizesse tal uso do computador, para tal propósito. Entretanto, o registro e a comunicação de programas e notas de programação poderiam ser feitos da maneira descrita para registrar e comunicar coisas como estas, escrevendo em folhas de papel, à mão, ou utilizando máquinas de escrever. Mas uma vez que o uso do próprio computador foi reconhecido para tais propósitos como uma possibilidade, e foi praticado regularmente, não é de se surpreender que surgissem algumas pessoas, com percepção suficiente para sugerir uma visão mais

ampla, e mais excitante, de suas possibilidades de uso, tornando real o que Vannevar Bush teria vislumbrado com o Memex, que seria, entre outras coisas, uma visão prototípica daquilo que mais tarde se tornaria o conceito de ligação entre hipertextos.

De qualquer forma, o próprio Skagestad apresenta três conclusões preliminares a partir de suas perspectivas históricas da diferença dessas duas visões:

> Primeiro, a máquina de Turing e o Memex, cada um destes forneceu um pedaço indispensável da tecnologia que acabou sendo conhecida como o computador pessoal, e que podemos hoje escolher por conceitualizar como uma máquina de Turing pessoal ou como um Memex computadorizado. Segundo, estas duas construções não são rivais, no sentido de oferecerem soluções conflitantes para o mesmo problema; Bush e Turing estavam abordando problemas inteiramente diferentes, e então suas respectivas soluções não conflituam diretamente.
>
> Terceiro, estas duas construções incorporam concepções diferentes da mente humana em geral e da interação homem-máquina, em particular.'

Ele continua, afirmando:

> Turing considerava o ser humano como essencialmente indistinguível de uma máquina; Bush considerava o ser humano essencialmente como sendo um usuário de máquinas, e procurou construir máquinas de manipulação de símbolos que seriam antes 'máquinas pensantes', no sentido de 'máquinas para se pensar com', e não 'máquinas que pensam'. Enquanto a visão de Bush serviu de inspiração para uma vasta indústria que está transformando rapidamente nossa cultura e sociedade, a visão de Turing tornou-se o paradigma diretor do programa de pesquisas conhecido como Inteligência Artificial, e também de toda a área interdisciplinar conhecida como Ciência Cognitiva. É tão presente a influência deste paradigma que freqüentemente ouve-se dizer que o único modelo de mente disponível, compreensível e bem detalhado, é o modelo computacional. Há, entretanto, um outro modelo de mente que se encontra disponível - um que, embora não tenha sido elaborado por Bush, dá total apoio ao programa de pesquisas que Bush iniciou, programa hoje conhecido como 'Inteligência Aumentada'. O modelo a que me refiro foi desenvolvido no século XIX por Charles S. Peirce, e foi recentemente defendido por James Fetzer como o modelo semiótico da mente.

Para sumarizar, o argumento de Skagestad equivale a dizer que as pesquisas em inteligência computacional (Pesquisas em IC) têm,

até agora, se desenvolvido principalmente em função de duas perspectivas diferentes, considerando os objetivos que podem ser alcançados - Inteligência Artificial e Inteligência Aumentada. Tais perspectivas podem ser vistas como alternativas complementares, não exclusivas, de modelos de desenvolvimento em IC, que, entretanto, podem entrar em desacordo, devido aos diferentes conceitos de mentalismo que as subsidiam. O objetivo primário de Skagestad não foi o de encorajar o desenvolvimento de pesquisas nas quais estas áreas pudessem mutuamente ser utilizadas como suporte, embora indubitavelmente ele fosse favorável a isso, mas deixar claro que o segundo paradigma de pesquisas em IC é conceitualmente independente do primeiro, uma vez que, aquilo a que nos referimos como sendo uma única coisa, o computador, corresponde, na verdade, a duas coisas: um mecanismo corporificador de algoritmos, capaz, até certo ponto, de imitar funções mentais, e um instrumento para a coordenação de diversos fatores envolvidos na inteligência humana, uma vez que podem ser mecanicamente realizados, de forma a aumentar a inteligência humana.

Skagestad considera a base teórica para a concepção da Inteligência Artificial como fundamentada na concepção de Turing de Máquina Universal, mas não considera o respectivo personagem histórico em Inteligência Aumentada, Vannevar Bush, como o fornecedor da base teórica para a tradição geral em Inteligência Aumentada. Sua visão, ao contrário, é a de que, embora Peirce não tenha vislumbrado essa área, como ela é hoje, de uma maneira concreta como Bush fez com sua conceitualização da máquina Memex, a filosofia de Peirce fornece uma base teórica para a tradição geral da Inteligência Aumentada, de um modo que a perspectiva mais limitada de Bush não seria capaz de fazer. Skagestad também reconhece outros personagens, cujas concepções dão suporte a essas bases teóricas, mais particularmente Karl Popper e sua concepção do desenvolvimento evolucionário exosomático da mente (Skagestad 1993). Mas ele considera o trabalho de Peirce, que é anterior ao de Popper, como sendo o mais adequado, do ponto de vista teórico.

Existe um princípio para a pesquisa em Inteligência Aumentada em geral?

Concordo com Skagestad, tanto em relação à necessidade de reconhecer que existem dois projetos de pesquisas distintos no de-

senvolvimento de tecnologias em inteligência computacional, quanto à afirmação de que a filosofia de Peirce pode fornecer uma base teórica para o segundo tipo de projeto em inteligência computacional, trazendo, ao mesmo tempo, contribuições importantes para o primeiro. Assumo isso como tácito, aqui. Mas antes de seguir e explicar os aspectos seguintes da tradição de pesquisas em Inteligência Aumentada que me interessam, devo primeiro mencionar que não acredito que Skagestad tenha sido bem sucedido, até agora, em identificar, de maneira suficientemente precisa, o que há de fundamental na tradição em Inteligência Aumentada, que vai de Bush, passando por Douglas Engelbart, J.L.C. Licklider (desenvolvimento da Internet), Ivan Sutherland (computação gráfica), Ted Nelson (hipertexto), Alan Kay (*design* de interfaces), e outros personagens, até Tim Berners-Lee. Estes, que criaram o conceito de *World Wide Web* e, ao mesmo tempo, o transformaram no sistema de hipertexto mundial, por volta de 1989, continuam, com seus trabalhos, no desenvolvimento da chamada 'rede semântica' (*semantic web*)[6]. Não encontro um lugar onde Skagestad descreva a Inteligência Aumentada de um modo que pareça capturar o que suas diversas facetas têm em comum, o que justificaria considerarmos esta segunda visão controladora, ela própria, como uma visão única ou unitária. Acredito, entretanto, que há realmente alguns fatores unificadores a serem considerados.

Assim, Skagestad refere-se algumas vezes à Inteligência Aumentada como sendo baseada na concepção de *computador pessoal*, em contraste com a concepção de computador que poderia ser exemplificada pelo tipo de computação característico da computação em *mainframe*s. Isto poderia talvez ser afirmado, identificando-se algumas peculiaridades características em computadores pessoais, das quais fosse possível derivar os princípios gerais encontrados na Inteligência Aumentada. Mas não creio que isto é feito satisfatoriamente. Ele também menciona a problemática e os propósitos do *design de interface* com o usuário como sendo de primeira importância, e isso, apesar de estar correto, também não é definido de maneira adequada. Usando a visão de Bush da máquina Memex como uma base histórica, ele está, na realidade, privilegiando os princípios do *hipertexto* como fundamentais, e isso certamente também é de suma importância. Mas, novamente, não vejo da parte de Skagestad nenhuma tentativa em demonstrar que estes princípios estão, de alguma forma, nos fundamentos de tudo isso. O conceito de *redes de computadores* poderia ser ainda um outro

candidato, que ele usa como ilustrativo da segunda revolução em computação, mas a idéia geral que está por trás das redes teria de ser colocada de maneira mais clara, além de mostrar que a mesma é conceitualmente básica, considerando os outros fatores mencionados como característicos da pesquisa em Inteligência Aumentada e isso não foi feito.

Meu palpite é que a chave para a identidade do que Skagestad caracteriza como a tradição 'Inteligência Aumentada' em pesquisas computacionais jaz na concepção de computação *interativa*, que ele de fato reconhece de passagem. Uma razão para se pensar que este pode ser o fator chave é que o conceito de computador pessoal parece ter se desenvolvido historicamente, principalmente a partir das tentativas das primeiras comunidades *hackers* no MIT em se aproveitar das máquinas DEC que chegaram para competir com os mainframes IBM, uma vez que respondiam melhor às necessidades dos programadores do que os monólitos que as precediam. Estas necessidades incluíam a necessidade de *jogar* — o grande fomento à criatividade nos desenvolvimentos em computadores, na minha opinião — e os jogos criados eram do tipo interativo, envolvendo a produção de texto, que deveria ser produzido pelo jogador e interpretado pelo computador, e também produzido pelo computador e interpretado pelo jogador, em uma seqüência contínua de respostas e contra-respostas que simulava a interatividade humana com coisas em seu ambiente, sob o contexto de uma estrutura investigativa que dá sentido a todo o processo. Refiro-me, claro, aos jogos do tipo 'aventura', em particular, que eram jogos de descoberta baseados em pistas fornecidas por descrições textuais sobre quais itens deveriam ser encontrados nos túneis labirínticos da 'Caverna Colossal'.

Com isso, o paradigma do computador, como uma máquina que executa algoritmos, perde o lugar para outra visão de como as coisas acontecem realmente. Isso porque, a despeito do que estava acontecendo no lado da *máquina* — assumindo que não há nada além do uso de algoritmos aplicados em estruturas de dados — o que estava acontecendo no lado do *jogador*, que também é parte integrante do sistema interativo global, não era algorítmico. Como resultado, o sistema interativo, como um todo, não poderia ser entendido simplesmente como o disparo ordenado de algoritmos, mantendo apenas uma pálida semelhança com aquilo que uma máquina parecia ser, considerando a visão de um programador de

mainframes, que estava acostumado a pensar em termos da máquina como dedicada à execução de rotinas puramente dedutivas operando sobre dados a ela fornecidos com o propósito de apenas extrair deles conclusões dedutivas. Encontrar o caminho de saída da 'Caverna Colossal' exigia muita dedução, mas dedução algorítmica não era exatamente a forma de atividade do sistema como um todo ('pessoa-e-máquina' interativo), pois, na realidade, a pessoa humaniza a máquina, dotando-a de espontaneidade humana a serviço da descoberta.

A interação entre seres humanos e máquinas na solução de problemas que surgem no contexto da descoberta — este é o ponto pelo qual eu começaria, na tentativa de obter uma visão clara e unitária da essência do que Skagestad considera como sendo a segunda revolução computacional e identifica como sendo o projeto da Inteligência Aumentada[7]. Skagestad certamente concordaria comigo neste ponto, não estou sugerindo nada discordante aqui. Mas o melhor que posso extrair daquilo que ele afirma nos artigos anteriormente mencionados, seu ponto de partida para entender a Inteligência Aumentada filosoficamente, é a idéia de uma localização 'exosomática' da mente no ambiente material. Deixe-me explicar agora como isto se relaciona com a afirmação Peirceana de que ‘todo pensamento é em signos’, que ele corretamente considera, em minha opinião, como a concepção chave para se entender a semiótica de Peirce como capaz de fornecer uma base teórica para a Inteligência Aumentada em geral.

O PENSAMENTO É EM SIGNOS — O PENSAMENTO É CORPORIFICADO EXOSOMATICAMENTE (SKAGESTAD)

Skagestad entende a afirmação 'Todo pensamento é em signos' como significando que o pensamento não é, primeiramente, uma modificação da consciência (uma vez que o pensamento inconsciente é possível na visão de Peirce) mas é, ao contrário, uma questão associada ao comportamento. Não é, entretanto, o caso do comportamento do ser que pensa, o ato de pensar (que seria aqui um caso especial), mas do comportamento do meio físico (que se encontra disponível publicamente) e dos artefatos nos quais o pensamento reside na forma de um poder natural. Este poder é a significação, ou seja, o poder do signo de gerar interpretantes. Pensamento é semiose, e semiose é ação do signo. Um signo se

realiza *como* tal ao gerar um interpretante, que por sua vez também é um signo subseqüente da mesma coisa, e que, ao ser realizado *como* um signo, gera também outro interpretante, e assim por diante.

Assim, como corretamente defende Skagestad, o desenvolvimento do pensamento acaba assumindo a forma do desenvolvimento do meio físico do pensamento, ou seja, coisas como o desenvolvimento dos instrumentos e dos veículos de expressão, tais como sistemas notacionais, ou meios e mídias de inscrição como livros e instrumentos de escrita, linguagens consideradas como entidades materiais como inscrições escritas e sons, instrumentos físicos de observação como tubos de ensaio, microscópios, aceleradores de partículas, e assim por diante. A evolução da mente significa que a cognição está se desenvolvendo, não fundamentalmente no sistema nervoso e no cérebro, não em algum tipo imaterial de coisa mental, mas ao invés disso nos instrumentos materiais e nas mídias da cognição. Assim,

> Um psicólogo remove um lóbulo do meu cérebro (*nihil animale a me alienum puto*) e então, quando vê que não posso me expressar, diz, 'Veja, sua faculdade da linguagem estava localizada naquele lóbulo'. Sem dúvida estava; e então, se ele tivesse surrupiado meu tinteiro, eu não estaria apto a continuar minha discussão até que conseguisse outro. É, [...], *os pensamentos não me viriam* [a ênfase é minha]. Assim, minha faculdade de discussão está igualmente localizada em meu tinteiro' (CP 7.366).

Deixem-me citar o comentário de Skagestad sobre isso:

> Como indicado pela sentença enfatizada, Peirce não está defendendo o ponto trivial de que sem tinta ele não estaria apto a comunicar seus pensamentos. O ponto é, ao invés disso, que seus pensamentos lhe vêm através do ato de escrever, de forma que implementos de escrita são uma espécie de condição para certos pensamentos — especificamente aqueles que surgem a partir de seqüências de pensamento que sejam muito longas para permanecerem na consciência. Esta é exatamente a idéia que, sessenta anos depois, motivou Engelbart a inventar novas tecnologias para escrever, de forma a melhorar o processo humano de pensamento, a idéia que motivou a interpretação que Havelock fez de Platão.

É facilmente perceptível a conexão disso tudo com o desenvolvimento da computação gráfica, da interface com o usuário, do uso do mouse, processadores de texto, hipertextos, e assim por dian-

te, que é o que fundamentalmente interessa a Skagestad. O embasamento teórico de tudo isso em Peirce reside no fato deste localizar o pensamento nos meios de sua expressão, o que fica expresso na afirmação de que 'todo pensamento é em signos'.

O PENSAMENTO É EM SIGNOS — O PENSAMENTO É DIALÓGICO (RANSDELL)

Permitam-me explicar uma coisa sobre meus próprios interesses na Inteligência Aumentada. Concordo com Skagestad, considerando o que foi dito até agora, e meus interesses certamente incluem os mecanismos que constituem e controlam a interface, tanto com documentos quanto com dados, quanto com pessoas, e que incluem ou permitem capacidades de manipulação de textos e gráficos que, desenvolvidos nos últimos anos, possibilitam criar e seguir links de hipertextos (i.e. fazer associações livremente e rastrear as associações já feitas), trocar mensagens com outros e nos comunicar com eles de diversas maneiras. Mas há uma interpretação posterior, igualmente válida, da afirmação de que 'todo pensamento é em signos', que também tem implicações para a Inteligência Aumentada baseada em computação, qual seja, a de que *o pensamento é dialógico* — e portanto, *comunicacional* — em sua forma. Se o pensamento deve ser encontrado em signos, e se este se consolida na geração concreta de signos interpretantes, então é o fluxo do discurso, visto como uma seqüência de interpretações assimétricas dialogicamente estruturadas, levando a interpretações consecutivas, que constitui o fluxo ou o processo do pensamento, e o desenvolvimento da inteligência é, no mínimo, uma questão sobre o desenvolvimento de *práticas de controle crítico* que se conformam a *normas comunicacionais* que tornam o discurso mais eficiente e efetivo, qualquer que seja sua finalidade.

Uma vez que se deseja que o discurso, ou a comunicação, seja efetivamente mais inteligente, parece ser razoável começar trabalhando com a comunicação, especialmente a forma como esta ocorre em processos de *investigação*, onde a função das normas de controle crítico é a de tornar a investigação mais bem sucedida. A habilidade de ser bem sucedido, desta maneira, é certamente uma parte importante daquilo que consideramos como inteligência e é, obviamente, um lugar natural por onde começar, para qualquer

filósofo que tenha sido influenciado pelo trabalho de C.S.Peirce e John Dewey, como eu fui.

Bem, até agora os tipos de investigação mais efetivos já criados pelo homem são aqueles que ocorrem em *tradições de pesquisa* como as que começaram a ser desenvolvidas na antiguidade, nos séculos VI ou VII a.C., e que têm se ramificado em muitas outras tradições, especialmente durante os últimos cinco ou seis séculos, incluindo o que agora chamamos de 'ciências' e também o que normalmente pensamos como tradições 'acadêmicas'. Nestas tradições, as técnicas de pesquisa estão incorporadas em práticas, hábitos, e competências dos investigadores que podem ser divididos em dois tipos: de um lado, as que podem ser chamadas de 'técnicas *materiais*' de investigação, ou seja, que foram desenvolvidas em um campo. (Algumas serão específicas de certos campos, muitas serão comuns a muitos campos, e algumas a todos os campos de pesquisa.) De outro lado, há aquelas que chamarei de 'técnicas *discursivas*' de investigação, significando com isso o domínio daquelas práticas, hábitos e competências de discussão e interação comunicacional que controlam o fluxo do discurso no contexto da investigação, de acordo com as normas desenvolvidas nas diversas tradições de pesquisa em geral: quero dizer, práticas especiais tais como a asserção, a sugestão, o questionamento, a elaboração de respostas críticas, a elaboração de contra-respostas, a capacidade de levantar objeções, o detalhamento de pontos levantados, etc. Estas habilidades têm sido pouco investigadas até então, e quero aqui trazê-las à sua atenção para tentar expressar algumas idéias sobre porquê as considero importantes, mesmo tendo sido largamente ignoradas como um tipo de pesquisa em Inteligência Aumentada, até o momento.

O tipo de Inteligência Aumentada em que especificamente me detenho é, portanto, aquele que se pode alcançar desenvolvendo mecanismos e programas para aumentar a efetividade das normas comunicacionais que proporcionam uma investigação bem sucedida, tais como os que se desenvolveram nas tradições de pesquisas, cujas formas ancestrais algumas vezes encontram-se há mais de dois milênios atrás, e também aqueles que facilitariam e investigariam as próprias normas, com o propósito de identificar aquelas com as quais uma conformidade iria, sem dúvida, produzir uma investigação mais bem sucedida. O projeto de desenvolvimento de qualquer dispositivo computacional que pudesse ser útil nesta em-

preitada poderia ser qualificado como uma contribuição à pesquisa em Inteligência Aumentada, desse tipo especial.

Devo destacar, no entanto, que se o foco sobre comunicação se detém na investigação, em particular, devemos nos questionar se esta é capaz de fornecer uma base adequada para entendermos as potencialidades da programação em Inteligência Aumentada, considerando que esta deve ser projetada para realizar uma comunicação mais inteligente. A abordagem que privilegia a investigação é um lugar natural para se começar, mas só pode nos conduzir até certo ponto, a partir do qual será preciso considerar outros tipos diferentes de comunicação igualmente importantes, caso nosso objetivo seja o de desenvolver uma Inteligência Aumentada que venha a ser a mais abrangente possível. Podemos deixar esta questão de lado aqui. Mas compreender alguma coisa sobre a potencialidade e sobre a problemática da Inteligência Aumentada a este respeito iria ao menos nos fornecer um entendimento mais sofisticado do papel das normas comunicacionais na vida intelectual, o que nos permitiria aproveitar o trabalho filosófico de Peirce, mestre em investigação em diversos campos, para desenvolver concepções analíticas com este propósito.

Investigação e asserção

O suporte para este tópico na filosofia de Peirce se encontra principalmente em sua teoria da investigação, que é o *framework* geral no qual ele se baseia para desenvolver sua lógica. Sua Lógica inclui o desenvolvimento de notações, técnicas de derivação para dedução, e metodologias de indução e abdução. Mas Peirce situa as questões lógicas tradicionais dentro de um *framework* de investigação concebido de tal maneira que poderia ser considerado, para alguns propósitos, como uma *teoria geral da asserção*. Entretanto, hesito em chamá-la assim pois isso poderia ser mais uma fonte de confusão do que algo útil, tendo em vista a maneira pela qual a teoria dos atos da fala, da qual Peirce foi um pioneiro, foi desenvolvida depois de sua morte. Esta vem se tornando uma abordagem diferente para se entender o que é uma asserção, ao minimizar, tanto quanto possível, o aspecto social do ato da fala. Isto é feito, considerando que o papel do destinatário no ato fica limitado àquilo que está implícito no reconhecimento de um ato. 'Pegar' é um termo usualmente utilizado para expressar este tipo de aquiescên-

cia constitutiva de um ato da fala como sendo deste ou daquele tipo, focando somente (normalmente com grande brevidade) no aspecto do envolvimento da comunidade, em geral, em todo ato sério de asserção, uma vez que o estudo do papel a que atos assertivos em particular se prestam, em uma *comunidade* de investigadores, tem sido largamente deixado inexplorado. Isto não é, entretanto, o que Peirce tinha em mente ao conceber a lógica como uma teoria geral da asserção.

O leitor familiarizado com o trabalho de Peirce sabe que ele fez, como prefácio de seu primeiro relato sistemático de compreensão da lógica da ciência, um par de ensaios — 'The fixation of belief' (CP 5.358), 'How to make our ideas clear' (CP 5.388) — que situam a lógica, em seu sentido mais estrito, aquele ensinado em aulas de lógica, dentro do *framework* geral de um processo de investigação[8] que poderia ser descrito assim: uma investigação particular, que ocorre dentro de um processo de investigação mais demorado não deve ser considerada como algo que tenha um início em um momento absoluto no tempo, nem em um ponto final, onde termina completa e definitivamente, mas deve ser idealizada como algo que chega à existência quando um processo em curso se torna informado por duas ou mais tendências conflituosas, com relação à *aceitação* de algo que surgiu de um empate ou impasse conceitual (*aporia*), do tipo que poderíamos descrever logicamente em termos de duas ou mais asserções de opinião contraditórias feitas simultaneamente. Uma investigação é constituída pela inabilidade dos investigadores para resolver um desacordo sobre o que deve ser aceito. Tal desacordo deve ter surgido como resultado da acumulação de entendimento até este ponto, e a direção global da investigação é dada pela tentativa de tomar os passos necessários para ultrapassar o impasse inicial ou *aporia*, a fim de se chegar a uma aceitação compartilhada e não-conflituosa dos resultados ou achados. Esta aceitação compartilhada, caso ocorra, irá permitir investigações futuras do mesmo tema, usando, quando for relevante, tudo o que puder então ser aceito como base para se chegar a um futuro entendimento sobre este tema. Os padrões típicos de acordo, desacordo, e estratégia de pesquisa que isso pode envolver, têm sido razoavelmente bem explorados, com relação a seu ponto de vista lógico, tanto por Peirce quanto por Dewey, e não serão aqui objetos de minha preocupação.

Agora, considerar a lógica como uma teoria da asserção é assumir uma perspectiva muito especial do processo de investigação[9], tomando-o do ponto de vista do investigador, que é considerado motivado como sendo membro de uma comunidade de pesquisa com o objetivo de fazer uma contribuição para o entendimento compartilhado do tema, ao que já foi desenvolvido pela tradição de pesquisa. O ato de asserção ocorre quando o investigador, tendo se preparado para assumir o risco envolvido em fazê-lo, tenta capturar a atenção de outros, em seu campo de pesquisa, de tal forma que eles acabem por chegar à mesma conclusão à qual ele chegou, e assim contribuir para a tradição de pesquisa, ao formatá-la na direção de um entendimento do assunto que seja, em último caso, compartilhado e estável.

Isto é feito a partir da afirmação de um *achado* ou, caso este seja considerado suficientemente importante, de uma *descoberta*, o que é feito a partir da publicação de um relato de pesquisa. Sua ocorrência, quando reconhecida, acaba por ser o disparo intencional de um conjunto complexo de obrigações e permissões comunicacionais, que se aplicam não somente ao pesquisador que faz a afirmação mas a todos aqueles da tradição de pesquisa envolvidos pela asserção.[10]

Asserção séria e publicação primária

Como veremos a seguir, é necessário distinguir entre asserções feitas de maneira séria, de outras, feitas jocosamente ou, pelo menos, não-seriamente. Inicialmente tratarei da asserção séria, tanto porque é mais fácil de caracterizar, quando comparada às muitas variedades de asserções não-sérias que tipicamente ocorrem ao longo de uma investigação (constituída por um certo número de diferentes maneiras, ao mesmo tempo importantes e por vezes sutis, nas quais a força de uma asserção pode ser qualificada), quanto por causa do papel único da asserção séria nas interações comunicacionais em curso, que estão continuamente estruturando e reestruturando o processo de investigação por meio do efeito de conformidade às normas de permissão e obrigação que estas envolvem. Considerado a partir de uma perspectiva um tanto quanto imparcial ou estética, o curso de investigação, em uma tradição de pesquisa que se encontra viva, exibe o que poderia ser visto como um tipo de coreografia da conversação, embora seus participantes

não pensem normalmente desta maneira. E, na 'dança da pesquisa', atos correspondentes a asserções sérias provêem um tipo de ênfase que tem um efeito de organização único no processo.

Para os presentes propósitos, deixe-me caracterizar uma asserção séria como a obtenção de que esta (considerando uma pessoa fazendo a asserção) deve assumir total responsabilidade por fazer uma afirmação que, uma vez considerada séria pelos outros na comunidade, colocará sobre eles a obrigação de considerar que o que foi reinvidicado é suficientemente sério para permitir que eles mesmos sejam persuadidos à conclusão a que o reclamante chegou, se o reclamante na verdade apresentou seus pontos de uma maneira que pode ser vista como racionalmente persuasiva. (Quem deve achá-la persuasiva? Na verdade, cada membro da referida comunidade de pesquisa, tomada distributivamente, i.e. cada membro tomado individualmente um a um, distintamente de um único indivíduo considerado coletivamente. A comunidade de pesquisa não deve ser considerada como uma entidade coletiva.[11]) Outras obrigações, envolvendo tanto o reclamante como seus colegas pesquisadores referenciados na afirmação, estão também envolvidas em uma asserção séria.

Por exemplo, exige-se do reclamante sinceridade sobre ter chegado à conclusão por si próprio; aqueles que são citados pela afirmação devem levar ao reclamante e à comunidade de pesquisa qualquer objeção séria que possam ter contra a afirmação feita, caso localizem alguma falha que considerem importante o suficiente para avisar aos demais membros da comunidade. A qualquer um citado pela afirmação — i.e. qualquer membro da comunidade de pesquisa — é permitido responder apropriadamente à afirmação, de qualquer maneira que considere adequada, desde que trate da questão sobre se a afirmação deve ou não ser aceita. Exige-se da pessoa que faz uma afirmativa que esta inclua informações suficientes sobre quais os métodos de replicação de resultados que poderiam ser utilizados para permitir que estes possam ser testados, segundo as especificações próprias do afirmante. Espera-se do afirmante uma explicação, caso alguma objeção seja feita com relação a tentativas falhas de replicação dos resultados, e assim por diante.

Isso descreve o que tenho chamado de asserção séria, que obviamente tem um papel especial no processo de investigação, devido

ao poder de afirmação que uma pesquisa feita seriamente tem, desde que seja considerada como tal por todos os envolvidos, de afetar o curso atual da pesquisa em uma dada comunidade, em virtude de suas habilidades para impor tais obrigações aos membros desta comunidade e, assim, algumas vezes obrigar os membros da comunidade em geral a uma conclusão em comum. Isto é o que o afirmante espera como efeito último ao fazer sua asserção, embora não haja nenhuma maneira de garantir que tal acordo seja atingido, de modo regular. De fato, o número daqueles que são bem sucedidos a este respeito são freqüentemente uma minoria. Não há nada como, por exemplo, um algoritmo que garanta a aceitação de um conjunto de pesquisas, e qualquer tipo de programação computacional que tome isso por objetivo seria fútil.[12]

Agora, uma asserção deste tipo corresponde, obviamente, àquilo que é normalmente chamado de 'publicação'. Mas a palavra 'publicação' é normalmente utilizada para se referir às diferentes maneiras de tornar pública alguma coisa, o que não implica ou traz em si esse tipo forte e definido de vinculação a normas, que está associado as afirmações de pesquisa, propriamente ditas. Iremos, então, nos referir a estas afirmações sérias de pesquisa como atos de *publicação primária*. (Um sinônimo adequado, neste contexto, poderia ser *publicação formal*, e irei realmente utilizá-lo algumas vezes. Mas há algumas razões para que um termo distinto seja utilizado, e, além disso, há uma motivação especial para adotarmos a palavra 'primária' para este propósito[13]).

Asserção não-séria

Mas o processo de investigação não envolve simplesmente seriedade, como anteriormente discutido, mas também envolve muita (de fato, muita) atividade comunicacional de um tipo preparatório, que também afeta seus resultados consideravelmente, mas o faz de maneira diferente, uma vez que o que é dito não é feito na forma de asserções sérias, e portanto não demanda as mesmas obrigações rígidas que um ato de publicação primária implica. (Isto não significa que nenhuma norma se aplica: todo ato de discurso envolve algum tipo de norma bem conhecida, e mesmo o discurso mais jocoso, em um contexto de investigação, é governado por normas.) Seriedade ou não-seriedade, neste sentido especial, não é uma questão de como as pessoas se sentem: pode-se, de uma

maneira não-séria, argumentar sobre diversos assuntos com grande paixão e intensidade de convicção, de acordo com as opiniões existentes em um dado momento, e ainda assim argumentar não-seriamente por todos os envolvidos que o que está sendo dito não deve ser tomado como uma invocação à aplicação das normas comunicacionais rígidas que estão normalmente identificadas com afirmações sérias em resultados de pesquisa. O que torna uma afirmação séria, em termos de relevância, é o reconhecimento e a aceitação *de facto* da intenção que as regras especiais de discurso obtém, o que inclui as obrigações e permissões concomitantes a uma afirmação de pesquisa. Isto não é assunto de como alguém se sente, mas da disposição para aceitar a aplicação das normas comunicacionais, de maneira especialmente rigorosa, que estão associadas a tais afirmações.[14]

Durante o desenvolvimento temporal das tradições de pesquisa, vários tipos de práticas comunicacionais foram desenvolvidos, entre as quais poderíamos encontrar algumas que qualificaríamos como não-sérias, no sentido indicado: por exemplo, discussões informais de natureza ocasional com colegas de pesquisa, incluindo correspondências por carta; grupos de discussão frouxamente estruturados de vários tipos, que podem abranger desde grupos de discussão locais, com tópicos mais ou menos definidos e agendas de discussão, até conferências internacionais, congressos e afins; esforços coordenados em grupos, como parte de projetos de pesquisa complexos tais como os que estão se tornando crescentemente comuns nas ciências exatas; mensagens postadas em fóruns públicos e *newsgroups*, e também *threads* de discussão que podem ser algumas vezes longas e complexas, até auto-comunicação, como quando estamos trabalhando nossas idéias, durante um isolamento momentâneo de indivíduos da mesma tradição com a qual nos identificamos.

Não tenho idéia de quantos tipos diferentes de práticas comunicacionais podem valer a pena reconhecer, mas eles irão, obviamente, variar imensamente, dependendo de como consideram as normas de controle que governam o que deve ser considerado comunicacionalmente apropriado, e também como se espera que a comunicação contribua para o objetivo geral de aprender mais, tanto em amplitude como em profundidade sobre o assunto-matéria da tradição de pesquisa. Algumas vezes, as pessoas precisam da oportunidade de experimentar novas idéias, simplesmente

para descobrir se vale a pena explorar mais adiante; outras vezes, há a necessidade de expor aos outros as idéias, para conseguir um rápido *feedback* crítico, seja ele negativo ou positivo; algumas vezes, as idéias são colocadas adiante, de tal forma a lançar as bases, que irão permitir futuras afirmações relacionadas ao pioneirismo da descoberta; outras vezes, certas coisas acabam sendo discutidas simplesmente porque os participantes assumem que a visão geral sobre os tópicos de pesquisa que os interessam estão demandando uma revitalização, e portanto devem ser colocados em um contexto fora do comum.

Quais desses casos seriam os mais importantes, considerando os objetivos de pesquisa? Seriam os casos de asserção séria (publicação primária) os mais importantes? A resposta é que, com certeza, não se pode fazer tal julgamento *a priori*, fora de um contexto de consideração, ou sem um entendimento de quanto a tradição de pesquisa em questão está florescendo, ou se está em um estágio onde não é claro para onde se desenvolve. Algumas vezes, uma afirmação de publicação primária pode ser de suma importância, e freqüentemente o é. Mas uma conversa casual de corredor entre um conjunto de pesquisadores de talento pode muito bem fazer uma grande diferença para o futuro da tradição de pesquisa, mais do que um ato de publicação pode fazer. Publicações primárias têm um papel único no processo que estaremos considerando mais adiante, mas 'importância' não é a palavra certa para isso. E isso deveria ser destacado, uma vez que há uma forte tendência não somente para super-enfatizar a importância de publicações primárias, mas concomitantemente para ignorar a possível importância de outras práticas comunicacionais, reduzindo o conceito do que é investigação a uma caricatura enganosa.

A atividade de pesquisa pode ser comparada a um tipo de caça. Tratar a publicação como a coisa mais importante na comunicação de pesquisa equivale a dizer que a coisa mais importante na caça é o ataque coordenado à presa o que é sem dúvida verdade em alguns casos, mas não pode ser dito ser verdadeiro de um modo geral, uma vez que um processo complexo como caçar pode muito bem envolver outras atividades que são somente preliminares à tentativa, no clímax do processo, de captura ou morte da presa. Essas atividades preliminares podem ser, na verdade, muito mais importantes para o sucesso da caça do que os atos finais de ataque e captura, que podem ser pouco mais do que *pro forma*.

A seguir, estarei ilustrando o que tenho em mente quando me refiro a este tipo especial de Inteligência Aumentada, chamando a atenção para um caso concreto de interesse não-usual — o sistema de publicação automatizado criado pelo físico Paul Ginsparg para benefício de sua própria comunidade de pesquisa (física teórica de alta energia), e para diversas outras associadas a ela. O interesse especial que podemos atribuir a este sistema se deve, em parte, ao fato de que para entendê-lo é necessário distinguir entre asserção séria e publicação primária, além de outros tipos de comunicação que ocorrem no fluxo da pesquisa. É importante ter em mente, entretanto, que não estou usando-o aqui como um paradigma geral de comunicação em pesquisa. Ele é aqui citado pela maneira como ilustra o papel especial que uma publicação primária pode ter na pesquisa, e também porque o conhecimento de como foi recebido por pessoas de várias comunidades de pesquisa interessadas no uso da Internet para comunicação científica e acadêmica revela a grande confusão que existe hoje no entendimento geral de como o controle crítico (baseado em críticas) funciona em comunicações de pesquisa. Esta confusão é decorrente, principalmente, de um mal entendido sobre a natureza e sobre a função da análise por pares.

O FUNCIONAMENTO DO SISTEMA DE PUBLICAÇÃO DE GINSPARG

Direcionemos nossa atenção para o caso do sistema servidor e de arquivamento automatizado para a distribuição de pré-impressões de publicações em física teórica de alta energia, e para diversos campos correlatos na física, astronomia, e matemática, desenvolvidos originalmente no Laboratório Nacional de Los Alamos, pelo físico Paul Ginsparg.[15] O sistema foi recentemente transferido para a Universidade Cornell, onde Ginsparg assumiu uma posição, e o nome oficial do sistema é simplesmente 'arXiv', onde o 'X' é um trocadilho visual sobre a letra Grega *chi*. Farei referência a ele como 'Sistema Ginsparg', a fim de evidenciar o trabalho de Ginsparg ao desenvolvê-lo, considerando-o uma aplicação em Inteligência Aumentada de especial interesse para nós. Uma vez que a investigação é uma forma de aprendizagem, cujo sucesso implica em um maior entendimento das coisas, qualquer coisa que contribua para a eficiência e efetividade da investigação é, *ipso facto*, um aumen-

to da inteligência. O interesse pelo trabalho de Ginsparg não está associado, entretanto, a qualquer novidade ou sofisticação especial envolvida na programação, considerada simplesmente como convencional, mas na maneira como — como um suporte material à comunicação, governado por normas de controle que, se espera, possam estimular o desenvolvimento da investigação nos campos a que originalmente pretendia servir. Estas normas são exatamente aquelas discutidas anteriormente, ou seja, as que governam aquilo a que me referi anteriormente como 'asserção séria' ou 'publicação primária'. O sistema Ginsparg foi obviamente preparado para ser um ponto de apoio a serviço deste propósito e, de fato, tem funcionado assim desde então.

A maneira como o sistema funciona é simples. Se algum pesquisador deseja fazer uma asserção, relacionada a um resultado de pesquisa, e a seus pares no campo de pesquisa em questão, ele escreve um artigo contendo a asserção feita (considerada como uma conclusão) e suas bases de sustentação, de tal forma a incluir qualquer coisa que possa ser necessária aos propósitos de teste ou replicação dos resultados, mesmo que estes envolvam apelo a uma razão *a priori*, tal como nos casos de provas matemáticas. Uma forma genérica para estes artigos poderia ser descrita de maneira razoavelmente específica, se necessário, mas, levando-se em conta nossos propósitos imediatos, não há a necessidade de ir além da afirmação de que não há nada de incomum na forma esperada para tais publicações, considerando as expectativas das pessoas nos campos que usam o sistema Ginsparg, como seu meio de publicação primária. Essa forma não difere significativamente da forma que tem uma publicação primária em qualquer outro campo de pesquisa. A maior parte destas observações pode ser deduzida do fato de que o pesquisador deve sempre deixar claro o que é necessário para a replicação de seus resultados. O sistema é programado para aceitar diversos formatos especiais de arquivos, tais como Postscript, PDF, LateX, e HTML. Espera-se que a pessoa que deposita o artigo faça a formatação e a codificação necessárias (ou garanta que estes assim estejam). O ato do depósito é visto pela comunidade de pesquisa como a proposição de uma asserção séria, ou seja, como um ato de publicação primária, e caso encontre certas condições mínimas (e.g. a inclusão de especificações para a replicação de resultados), é reconhecido como tal.[16]

Além do artigo, também se prepara um resumo (*abstract*) que o acompanha, normalmente envolvendo o uso de palavras-chaves referentes aos tópicos abordados, depositando-se tanto os artigos como o resumo no sistema de arquivos. O resumo (não o artigo) é então automaticamente distribuído por *e-mail* para todos os usuários do sistema de publicação que tenham indicado previamente, por meio de uma descrição de seus interesses individuais de pesquisa, que estariam interessados em ler os artigos contendo material pertinente a suas atividades de pesquisa. (Uma vez que o sistema de arquivos está dividido em campos e sub-campos, pode-se simplesmente indicar para a máquina que está interessado em qualquer resumo depositado em um desses campos). Se um leitor de um resumo decide que o artigo pode ser interessante, então ele pode clicar em um *link*, o que fará com que todo o artigo seja enviado, ou iniciará o *download* do artigo. Todo o processo de depósito, notificação, e recuperação de artigos é automático.

Caso alguém não concorde com alguma asserção feita no sistema, e considere este desacordo importante o suficiente para ser formalizado, pode depositar uma resposta no mesmo local, o que será considerado como formalmente correto. Assim, intercâmbios na forma de um diálogo crítico podem ocorrer por meio do sistema, equivalentes aos que ocorrem em revistas profissionais tradicionais que permitem respostas (*replies*) como parte do processo de publicação. Mas é importante ressaltar que a estrutura do sistema não é apropriada para discussões informais típicas de, por exemplo, um fórum de discussões baseado em listas (*listserver*), ou grupos de discussões organizados, ou entre membros participantes de um projeto, *bulletin boards*, *newsgroups*, e muito menos do tipo de discussão que pode ocorrer em um chat em tempo real. É possível que respostas inapropriadas possam ser feitas e depositadas no arquivo (não há nada que impeça isso) mas o sistema foi projetado de forma a desencorajar este tipo de uso, exigindo também o depósito de um resumo (*abstract*) junto da réplica, de modo a informar aos demais membros da área que uma réplica foi efetuada. Isso ajuda a assegurar, na prática, um tipo de formalidade que está na essência do que chamo de publicação primária. Há muita coisa em jogo, do ponto de vista profissional, naquilo que surge a partir da tentativa de se tornar apropriados os adendos comunicacionais de algo como, digamos, uma discussão informal em grupo. No caso do sistema servidor de arquivos Ginsparg, não

há nenhuma política especial para se garantir isso, da mesma forma como, na prática, isso não tem se mostrado necessário.

A importância obscurecida do Sistema Ginsparg

De uma perspectiva mais restrita, o sistema servidor de arquivos Ginsparg não é nada mais do que uma forma automatizada de um sistema de comunicação que existe há décadas em diversos campos de pesquisa. Isto é, trata-se da prática de distribuir cópias de pré-impressões de artigos para outros membros de uma mesma área de pesquisa, entendendo-se aqui 'pré-impressões' como artigos que incorporam asserções de publicação primária, distribuídos entre pares de pesquisa, antes que estes apareçam como artigos publicados em revistas ou periódicos controlados editorialmente na área, muitas vezes antes mesmo da submissão de tais artigos a estes meios de publicação, e algumas vezes nem mesmo tendo em mente tal submissão. Pré-impressões não são, entretanto, meramente esboços ou rascunhos, uma vez que isso implicaria em uma falta de polimento e/ou completude que não seriam apropriados em algo distribuído como pré-impressão. Por outro lado, uma pré-impressão pode ser vista como uma versão passível de revisão, e a maioria das pré-impressões que são encaminhadas para publicação, em algum periódico, irá provavelmente sofrer alguma revisão antes de sua aparição no periódico, mesmo que seja a pedido de seu editor, que está freqüentemente sob pressão para economizar o espaço destinado aos artigos.

Antes do estabelecimento do sistema Ginsparg em Los Alamos, a distribuição das pré-impressões significava uma distribuição para aqueles que estivessem suficientemente bem relacionados profissionalmente, de forma que pudessem estar nas *mailing lists*, gerenciadas pelos que estavam na 'vanguarda' da área. Isso, obviamente, garantiria uma grande vantagem no sucesso profissional. Desse modo, havia na verdade dois locais distintos de publicação primária nestas áreas: o sistema de distribuição de pré-impressões e o sistema de periódicos profissionais controlados editorialmente e revisado por pares (*peer-reviewed*), gerando uma distinção entre pesquisadores bem-relacionados, e com maiores vantagens, e aqueles não-tão-bem-relacionados, excluídos das posições para participar da pesquisa de vanguarda. O atraso de tempo envolvido na publicação de periódicos profissionais usualmente acaba por significar que,

quando aqueles que dependem da literatura dos periódicos para entender o que está na vanguarda percebem o que está acontecendo ali, as fronteiras da vanguarda já terão se deslocado. Qualquer área que coloque grande ênfase na prioridade de descobertas acabará por recorrer à distribuição de pré-impressões como um meio de publicação primária, a menos que exista algo que dificulte isso, e a dominação do direcionamento nas pesquisas em muitos campos, por aqueles que estão em posições privilegiadas uma vez que estão aptos a participar em publicações primárias deste tipo (o que algumas vezes era discutido em termos da dominação da pesquisa por 'colegiados invisíveis'[17] de privilegiados comunicacionalmente) era um assunto de preocupação crescente nas ciências, quando Ginsparg estabeleceu seu sistema servidor de pré-impressões automatizado e sem restrições de acesso em Los Alamos.

Ginsparg e seus associados estavam conscientes, desde o começo, de que algo de importância potencialmente singular havia sido conseguido pelo ato relativamente simples de se instalar um sistema servidor de arquivos na Internet com uma política de acesso sem restrições para depósito e recuperação. A coisa mais importante para eles parece ter sido que, ao adotar este novo sistema, estavam fazendo uma transição entre um sistema de publicações que servia principalmente aos interesses especiais de somente alguns físicos que, como eles mesmos, tinham a sorte de estar no pequeno grupo fechado dos mais avantajados, para um sistema capaz de servir às necessidades de todos os físicos do mundo que fossem capazes de acessar a Internet, mesmo que com somente um nível mínimo de eficiência, sem as limitações baseadas na necessidade de uma qualificação especial ou relacionamento com certos colegiados. Farei referência a isto como o motivo *cosmopolita*.

Ao mesmo tempo parecem ter entendido que algo mais estava acontecendo, e que isso tinha a ver com o fato de que ali se mostrava que as práticas de análise por pares utilizadas pelos periódicos não eram pertinentes (ou seja, não se aplicavam) ao controle crítico da pesquisa de vanguarda. Uma vez que é parte da sabedoria convencional o fato de que é justamente a análise por pares que garante que os 'padrões de qualidade' possam ser reconhecidos na pesquisa e no controle de publicações, a sua dispensa, sendo ela tipicamente desdenhada como não-pertinente, foi entendida como algo perigosamente subversivo à ciência e à academia, especialmente considerando-se o fato de que as disciplinas científicas das

quais esta prática estava emanando são muito bem qualificadas na escala de prestígio profissional e assim não se poderia simplesmente dizer que estas são queixas que se poderia esperar de pessoas que não tem a capacidade para conhecer os supostos altos padrões da análise por pares. Este pode ser considerado como o aspecto *anti-autoritário* do idealismo deste grupo porque é de fato uma rejeição à concepção autoritária do papel da análise por pares na pesquisa, e penso que eles tiveram algum entendimento disto, mesmo que não veja aí uma tentativa de se re-pensar o conceito de análise por pares, para entender exatamente o que está ou não acontecendo, e qual é, e qual deveria ser, na verdade, a base para o controle crítico.

Desta forma, Ginsparg e seus associados assumiram uma visão altamente idealista deste fato, pelas razões anteriormente indicadas, e este zelo idealístico acabou assumindo a forma de se propor que o que eles tinham conseguido em Los Alamos para suas próprias áreas poderia ser aplicado nos demais ramos da ciência, e não somente ali mas nas tradições gerais de pesquisa. Limitações de tempo e espaço não permitem uma descrição aqui do que aconteceu depois, quando este entusiasmo encontrou-se com uma resistência crescentemente endurecida, que finalmente acabou assumindo a forma de uma retórica deflacionária, e que tem tido bastante sucesso para induzir um tipo de confusão obscurantista sobre o sistema de publicação Ginsparg, o que acabou por silenciá-lo como um movimento reformista.[18] Este objetivo foi atingido promulgando-se um certo mal-entendido, bastante importante, sobre a natureza da análise por pares, enquanto que, ao mesmo tempo, se proibia a discussão da reforma do processo de análise por pares nos fóruns públicos mais influentes, onde o tópico da educação *on-line* gratuita era discutido. Isso efetivamente reduziu a significação do sucesso deste sistema de publicação a um mínimo, uma vez que se encorajava uma recusa ao reconhecimento do sistema Ginsparg como um sistema de publicação primária.

Quando a existência do sistema Ginsparg tornou-se largamente conhecida, anos atrás, gerou-se uma visão extremamente alarmista, além de predições calamitosas sobre o declínio inevitável na qualidade de pesquisa nos campos onde o uso do sistema era comum.[19] Parece razoavelmente claro agora, entretanto, que este declínio previsto não ocorreu e as avaliações pessimistas parecem ter dado lugar a uma admissão, por vezes rancorosa, de que o

sistema parece funcionar nos campos para os quais foi originalmente projetado. Por outro lado, tornou-se também cada vez mais claro que não há ainda uma tendência com relação a sua adoção como um modelo geral de prática de publicação em ciências, como Ginsparg e alguns de seus associados haviam imaginado que poderia ocorrer, muito menos no sentido de sua emulação em publicações de pesquisa científica e acadêmica de maneira generalizada. Conseqüentemente, o interesse inicial neste, como um sistema novo e revolucionário viabilizado pela Internet, acabou por desaparecer.

Assim, como indicado anteriormente, o sistema foi considerado como não sendo um sistema de publicação, apesar de ter continuado a ser o principal sistema para publicação primária, como definido aqui, nos campos para os quais foi concebido. Desta maneira, seu único valor em relação às práticas de publicação em geral, conforme usualmente se considera, reside no fato de que ele acabou por fornecer um modelo para o desenvolvimento de sistemas de arquivo de Internet, de um tipo que pode ser copiado em qualquer número de diferentes nós na Internet — sistemas de arquivos deste tipo, localizados em universidades, estão agora sendo apresentados como réplicas localizadas e perfeitas dele. Sua virtude vem do fato de que qualquer coisa depositada em qualquer um destes sistemas de arquivos torna-se disponível como um documento em uma única base de dados virtual, em qualquer lugar do mundo. Seus documentos são suscetíveis a buscas, além de ficarem disponíveis a programas desenvolvidos com o propósito de extrair material dele, de modo a acompanhar o que existe ali, como qualquer bibliotecário poderia fazer, e também com o propósito de analisar os documentos ali contidos de forma a separá-los por tipo e a descrevê-los, de acordo com qualquer tipo de critério, correspondendo aos vários interesses que alguém poderia ter neles. O valor disto é inquestionável, mas esta não é, na minha opinião, a coisa mais importante a se entender sobre o sistema de publicação Ginsparg.

Assim, embora a desinformação retórica sobre o sistema, como um sistema de publicação, não tenha tido nenhum efeito sobre seu uso nas áreas para as quais foi projetado, onde ele ainda continua a prosperar, houve de fato um desvio de foco sobre seu aspecto mais idealístico e sobre a potencialidade para encorajar reformas que estavam implícitas no sistema automatizado. Sua relevância acabou parecendo ser (um tanto erroneamente) somente a de um

exemplo de como seria possível fazer a transição de periódicos, baseados em papel, para publicações *on-line*. Isso, desconsiderando qualquer tipo de reforma que poderia atrapalhar os sistemas já instaurados de hegemonia exercidos pelas diversas instituições e grupos que controlam a pesquisa, dando suporte e controle aos mecanismos de publicação. Com isto, a significação do sucesso do sistema de arquivos Ginsparg, e sua contribuição para o desenvolvimento do que é potencialmente uma parte muito importante da pesquisa em *Inteligência Aumentada* têm sido obscurecidas de tal forma a serem comparadas a um tipo de 'emburrecimento' de nosso entendimento sobre as condições necessárias para o sucesso na pesquisa científica e acadêmica. Para reverter isto, é necessário insistir sobre o desafio que o sistema Ginsparg apresentou e continua a apresentar à análise por pares, como se entende hoje.

Análise por pares e o conceito de par

Devo enfatizar que a visão de análise por pares aqui proposta não deixa de considerá-la como de importância fundamental no controle crítico da pesquisa. A questão é que aquilo que passou a ser chamado de 'análise por pares' não é exatamente uma análise por pares, mas uma forma degenerada desta que é, como princípio de controle crítico, não somente de valor limitado, na melhor das hipóteses, mas também uma subversão do próprio princípio do que é um par, o que deveria estar por trás da prática de uma autêntica análise por pares. Por quê? Porque ela trata da análise por pares como um sistema de controle por parte das elites, o que é contraditório com o conceito de par. De acordo com a visão aqui apresentada, o funcionamento de uma análise por pares de fato autêntica, poderia ser melhor observado, em ação, pelo estudo das práticas paradigmaticamente exemplificadas pelo sistema Ginsparg (ou qualquer sistema equivalente) de publicação primária.

Quando fiquei interessado, pela primeira vez, por este assunto pensei que seria melhor não contrariar o uso atual do termo 'análise por pares' como se referindo à análise de pré-publicações por pares editorialmente comissionados, especialmente considerando que os primeiros entusiastas do sistema de publicação Ginsparg achavam a análise por pares, neste sentido, como de pouca importância real, uma vez que a pesquisa de vanguarda parecia fazer pouco uso disso: quando um artigo aparece num periódico, a vanguarda já

avançou e então a função do periódico como um espaço de divulgação não pode ser a de controlar o que aparece na vanguarda, em termos de asserções de pesquisa. Isto não equivale a dizer que um periódico editorialmente controlado não poderia estar a serviço daqueles que estão na vanguarda, mas isto teria de ser decorrente do uso da retrospecção, de ser claro sobre o que foi alcançado, diferentemente de seu mero alcance, que não exige uma validação editorial baseada em uma análise por pares. Algo assim foi, e talvez ainda seja, a visão dos pesquisadores que apóiam o uso do sistema de publicação Ginsparg.

A possibilidade de que a função de controle primário esperada na análise por pares fosse a de controle da vanguarda, e que o sistema automatizado de arquivos sem filtragens fosse o lugar onde se pudesse encontrar uma autêntica análise por pares, não parece ter ocorrido aos físicos, no entanto, e muito menos ocorreu aos partidários dos periódicos editorialmente controlados, que consideram o periódico como o local natural para uma análise por pares, bem como o fator de controle crítico primário na pesquisa científica. Com este tipo de acordo entre os antagonistas, relativo ao que seja uma análise por pares, ou seja, uma avaliação de pré-publicações por pares editorialmente comissionados, me pareceu que seria insensato lutar contra isto, mesmo que eu achasse que existia aí um grande engano. Assim o melhor seria deixar o uso da noção 'análise por pares' da maneira como estava. Afinal, tudo não passaria de uma questão verbal, ou algo assim, como inicialmente eu supus, e talvez não mais do que uma discussão sobre o uso do termo.

Mas eu estava errado. O que eu ainda não havia percebido é que foi tirando proveito do descaso dos físicos pela análise por pares, da forma como eles a entendiam em comum com seus oponentes, como um fator primário de controle crítico, que o radicalismo supostamente perigoso do sistema de publicações automatizado sem filtragem e sem restrições de acesso de Los Alamos seria neutralizado, tornando-se inócuo, considerando qualquer interesse institucionalizado que pudesse ser ameaçado pelo desafio colocado à análise por pares. Percebi, entretanto, que uma vez que é o respeito ao princípio da idéia de par que está nas bases do controle crítico de comunicações de pesquisa, que teria sido um erro retórico o que tem permitido o sucesso daqueles que negam o significado do sucesso do sistema Ginsparg, negando que ele tenha o *status* que

realmente tem, como um espaço de publicação primária. Com este *status* negado, o que quer que ocorresse no sistema Ginsparg poderia ser e agora comumente é, de fato, preterido como sendo não diferente em forma ao que acontece em qualquer outro boletim, fórum ou grupo de discussão baseado em listas, *chat* ou qualquer outro meio informal visto como não importante o suficiente para ser um desafio à hegemonia da legitimidade reivindicada pelos periódicos editorialmente controlados.

O que está faltando então é um entendimento melhor do conceito de par, o que explicaria porque a análise por pares, onde quer que esta esteja localizada no processo, pode ser vista como exercendo a função básica de controle crítico na pesquisa. E é exatamente por esta razão que o que hoje é normalmente considerado como sua ocorrência paradigmática — a análise de pré-publicações por pares editorialmente comissionados — deveria ser reconhecido como um falso postulante ao título, uma vez que uma análise por pares, devidamente entendida, seria melhor exemplificada hoje pelos trabalhos rotineiros do sistemas de arquivos e servidores automatizados, sem filtragem, sem edição, e sem restrições de acesso como é o sistema Ginsparg. Ou então algo como ele, eu argumentaria. Por que, então, considera-se importante que a aceitação de resultados de pesquisa em um dado campo seja algo que acontece em conseqüência da avaliação destes resultados por *pares*?

Um par de pesquisa é, naturalmente, um igual. Para ser mais exato, um par é um igual *de maneira presumida*, não alguém que tenha demonstrado ser de fato igual, neste ou naquele aspecto, mas alguém que possa ser considerado, presumidamente, como alguém cuja opinião informada sobre o assunto-tema de pesquisa devesse ser tomada tão seriamente como se fosse nossa própria opinião, assumindo que esta depende do *status* do pesquisador, ao contrário de uma possível dependência de uma *justificação dada* pelo pesquisador em suas asserções. Um par é alguém cujo desacordo com nossas próprias opiniões requer uma explicação e não simplesmente algo a ser ignorado. Isto deve ser assumido antes de qualquer especulação sobre a razão deste desacordo. Um não-par é alguém cuja opinião sobre o assunto em questão não nos faz a menor diferença, não nos coloca nenhum desafio, nem para que tentemos explicar esse desacordo, nem para que aceitemos o fato de que não soubemos expressar adequadamente nosso ponto de vista, pois não conseguimos nos explicar. A questão sobre o *status*

de par de uma pessoa não será discutida aqui. Podemos entender sua importância, mas não é possível fazer justiça a esta questão aqui. Basta-nos, para os presentes propósitos, dizer que, na prática, o que determina o *status* de par é o que determina a quem estamos nos dirigindo em nossas asserções de pesquisa.

Este é um dos muitos lugares onde o conceito de publicação primária mostra-se valioso. Em uma publicação primária, o pesquisador que faz uma asserção está sempre se dirigindo a uma comunidade de pesquisa mais ou menos definida. Esta consiste em todas as pessoas que compartilham da mesma preocupação com um assunto-tema. O objetivo de se fazer a asserção é o de apelar a tais pessoas para que elas reconheçam a validade da conclusão que foi obtida, e que está sendo apresentada como uma conclusão que, conforme se argumenta, pode ser aceita a partir deste instante, como premissa ou pressuposição, em seus pensamentos subseqüentes sobre o assunto-tema em questão. Ao se apresentar asserções na forma de uma publicação primária, tenta-se iniciar um processo controlado pelo complexo conjunto de obrigações e permissões a que aludi anteriormente e descrevi sem maiores detalhes, acima, o que se espera, possa resultar em sua aceitação eventual por parte daqueles a quem as asserções foram direcionadas. Esses, a quem as asserções foram direcionadas, são os pares (ser par de alguém, e ser membro do grupo para onde as asserções foram direcionadas, é a mesma coisa) e assim todas estas pessoas são consideradas como sendo iguais no sentido de estarem sujeitas às mesmas normas, cujo conjunto também se aplica ao próprio pesquisador fazendo as asserções, uma vez que estas são as regras definidoras de papéis às quais o diálogo da publicação deve se conformar.

Assim, a análise por pares propriamente dita, é o que ocorre no processo de investigação quando se faz tal asserção de pesquisa e a comunidade de pesquisa a quem ela é dirigida responde de acordo com as normas comunicacionais estabelecidas. Toda comunicação que ocorra dentro deste espaço dialógico constituído normativamente, e que pertença à asserção em questão, é uma análise por pares.

A RAZÃO DO PRINCÍPIO DE PAR NO CONTROLE CRÍTICO

Para um entendimento mais claro do que é a análise por pares e porque esta é corretamente considerada como fundamental no con-

trole crítico da pesquisa, temos que entender porque se julga importante que a aceitação das asserções de pesquisa, em um dado campo, seja algo que acontece em decorrência da avaliação destas asserções feita pelos pares. Como já destacado, um par é um igual presumido, não alguém que tenha demonstrado ser de fato igual, neste ou naquele aspecto, mas alguém que é considerado, presumivelmente, como alguém cuja opinião informada sobre o assunto-tema é importante o suficiente para que qualquer desacordo das opiniões produza uma situação na qual ambas não possam ser verdadeiras, mas não se possa decidir a respeito de qual está errada, argumentando-se que um deles tenha *status* superior. Em outras palavras, não pode haver uma relação de autoridade entre pares, a menos que estejamos falando sobre uma certa *fazenda de animais*[20], onde alguns pares são reconhecidos como sendo mais 'pares' do que outros.

O conceito de par aparece em diversos contextos diferentes da sociedade moderna. Um exemplo familiar sobre como funciona o *status* de par pode ser ilustrado por um caso onde um médico é chamado por um paciente de outro médico, para a emissão de uma segunda opinião. Médicos normalmente não se opõem a um pedido de um paciente por uma segunda opinião. Isso é entendido como se as duas opiniões estivessem em paridade, como avaliações profissionais, no sentido em que a segunda opinião é simplesmente mais uma opinião a ser plenamente considerada, ao invés de ser uma opinião definitiva ou determinante em relação à primeira: não há e não pode haver presunção geral a favor da opinião de um par relativa a de outro baseada na importância do médico. Eles são, neste sentido, iguais. Isto não implica que um deles não possa descrever o caso melhor do que o outro, mas isso é algo que o paciente terá que julgar por si próprio. No caso de duas opiniões conflitantes, a questão sobre qual deve ser seguida não pode ser estabelecida consultando um terceiro médico, que pudesse definir o assunto dizendo ao paciente qual está certa. Tudo que o terceiro médico pode fazer é oferecer uma terceira opinião, paralela às outras duas, e se concorda com uma e não com a outra; ainda assim não há implicação de que uma opinião aceita por dois dos três médicos será, em virtude disto, a melhor opinião. Em outras palavras, não há posição de autoridade reconhecida entre médicos. Em geral, não há autoridades entre pares, superiores ou inferiores. O reconhecimento da importância do par é assunto procedimental.

Se as coisas são assim, a questão a ser considerada é: por que a concepção igualitária de um par de pesquisa deveria ser considerada como uma parte do mecanismo racional normativo da pesquisa, como tem sido concebido nos tempos modernos? Isto é devido meramente a uma extrapolação solidária de um comprometimento com uma igualdade política? Mesmo que este tópico não possa ser aqui aprofundado em detalhes e com o rigor que merece, pelo menos se pode dizer que a adoção desta concepção normativa na prática de investigação está baseada na premissa de que, diante de uma interação perceptiva com o assunto-tema, na *experiência* dele, o próprio assunto-tema nos conduziria à crença ou convicção sobre ele, uma vez que nos fizéssemos adequadamente receptíveis a ele, conceitualmente e perceptivamente. A premissa é a de que deve haver uma relação causal entre o assunto-tema de pesquisa e o pesquisador, na qual o pesquisador é passivo no sentido de receber a ação do objeto, tal que as convicções do pesquisador sejam moldadas pelo próprio assunto-tema.

Uma ilustração do senso comum: que cor tem um certo objeto que está fora do meu campo de visão? Eu sigo passos para observá-lo e quando faço isso vejo que é vermelho, digamos, e não há duvida sobre isto. Posso pensar qualquer coisa que queira, mas o próprio objeto insiste em impressionar sua vermelhidão em mim, nesta situação particular, quer eu queira ou não. A experiência é o que a interação com o objeto imprime sobre você; é o que emerge do seu encontro com o objeto, como sendo apreendido a partir dele. Agora, é o princípio empiricista que requer o reconhecimento do princípio do par. Permita-me explicar rapidamente como isto funciona.

Coisas reais são multi-facetadas, no sentido em que podem ser percebidas a partir de múltiplos pontos de vista complementares, cada um deles sendo uma faceta ou aspecto da aparência da mesma coisa. Como o observador varia em sua relação com o objeto, a mudança na perspectiva ou no ponto de vista revela outras facetas do objeto, sendo que cada uma delas deve ser levada plenamente em consideração e conciliada com as demais, como diferentes facetas ou aspectos de uma mesma coisa. A razão por que devemos respeitar outros como nossos pares na investigação das coisas é que não podemos possivelmente construir um entendimento geral adequado de nosso assunto-tema, em um campo de pesquisa, sem confiar na competência básica de outros pesquisadores do campo,

exceto quando nós temos razões definitivas para duvidar deles, dado que temos alguma razão *prima facie* para supor que esta competência exista. Do contrário, nossa atenção teria que ser constantemente desviada para uma investigação da competência de cada um de nossos colegas, ao invés do assunto-tema. Um par é, logicamente considerado, equivalente a uma perspectiva respeitável (ou um conjunto de perspectivas) sobre o assunto-tema, e tratar um par como algo diferente de um equivalente, seja como superior ou inferior, é debilitar e perturbar a coordenação de perspectivas que é tarefa constante da ciência em desenvolvimento.[21]

Peirce, em uma passagem notável de 'How to make our ideas clear', descreve a coordenação das perspectivas dos pesquisadores individuais, onde se assume um igual respeito por cada uma destas perspectivas, cada uma tendo seu próprio papel na composição dos dados sendo reconciliados durante a coordenação:

> Um homem pode investigar a velocidade da luz pelo estudo do trânsito de Vênus e as anomalias das estrelas; outro pelas oposições de Marte e os eclipses dos satélites de Júpiter; um terceiro pelo método de Fizeau; um quarto pelo de Foucault; um quinto pelos movimentos das curvas de Lissajoux; um sexto, sétimo, oitavo e nono, podem seguir diferentes métodos de comparação das medidas de eletricidade estática e dinâmica. Eles podem, em princípio, obter diferentes resultados, mas, à medida que cada um deles aperfeiçoa seu método e processo, vê-se que os resultados movem-se continuamente em direção a um centro definido. Isto ocorre com toda pesquisa científica. Mentes diferentes podem partir das visões mais antagônicas, mas o progresso de investigação as carrega, através de uma força externa a elas, para uma mesma e única conclusão (CP 5.407).

A força externa a todos eles como indivíduos é a manifestação da realidade do objeto como agente causal, determinando o entendimento da comunidade de pesquisadores ao compelir as opiniões inicialmente distintas de cada pesquisador, de maneira a contribuir para o entendimento coletivo acumulado no qual, pela reconciliação e coordenação teórica, um único assunto-tema comumente percebido e compartilhado é atingido. Quando somente alguns membros de uma comunidade de pesquisa estão sendo tratados como tendo o direito de fornecer informações para a reconciliação teórica que está constantemente sendo construída, no curso da investigação, a comunidade de pesquisadores reduz-se ao número daqueles mais privilegiados. As propriedades do assunto-tema que estão sendo

acessadas e tomadas em consideração, com o propósito de se chegar a um entendimento do assunto-tema, estão correspondentemente diminuindo, e com isso, o entendimento teórico sendo erguido nestas condições está sendo cada vez menos determinado pelo real e, portanto, aumentando a chance de ser meramente fantasioso, uma vez que algumas de suas propriedades que poderiam estar sendo fornecidas pela comunidade de investigação não se encontram experiencialmente acessíveis. O princípio empírico de que o assunto-tema deve ter o seu algo-a-dizer nos resultados a que se chega, considerando-se que esses resultados sejam confiáveis, exige a aceitação do princípio de par na pesquisa.

De forma geral, é possível produzir, dentro de uma comunidade ou tradição de pesquisa, um entendimento de maior ordem de magnitude de inteligência do que seria alcançável por um indivíduo. É como se cada pessoa na comunidade de pesquisa, cada um vendo algo que outros não percebem exatamente da mesma forma, acrescentasse sua faculdade de percepção individual e tudo o que pode dela advir ao que é produzido por seus pares, colocando-se desta maneira, a serviço dos outros. A comunidade de pesquisa desta maneira tem tantos olhos, ouvidos e mentes individuais quanto tem membros confiáveis, e é incomparavelmente mais inteligente do que qualquer um de seus membros individualmente considerados, dado que suas práticas comunicacionais permitem que seus membros trabalhem em conjunto na construção de um entendimento teórico confiável que possa ser comumente aceito e no qual as contribuições individuais são reconciliadas e coordenadas.

Isto pode ser feito, então, uma vez que haja relações básicas de confiança na competência e honestidade entre os membros da comunidade de pesquisa considerada, e esta confiança deve ser, por outro lado, baseada em atitudes de presunção mútua, que eliminam a necessidade de que cada pessoa tenha sua confiabilidade constantemente atestada pela visão dos outros. Tente imaginar o caso de um pesquisador em uma comunidade de pesquisa que suponha que as opiniões de todos os demais membros da comunidade de pesquisadores só possam ser consideradas aceitáveis depois de terem sido verificadas ou corroboradas por ele mesmo. Sem uma presunção de competência e integridade em relação aos outros, na condição de reconhecer cada um dos outros como um par, não se pode falar em comunidade, nem tampouco há uma visão compreensível disponível sobre o assunto-tema. Esta condição de reconhe-

cimento de pares se reproduz a cada atualização das normas que regulam o processo comunicacional, muitas das quais poderiam ser encontradas por meio de uma análise (eu poderia assim argumentar, mas me restringirei a sugerir), o que deveria ser baseado no princípio do par, fornecendo o *framework* de obrigações e permissões presumidas que permitem a coordenação da opinião para obter e sustentar a si própria ao longo do tempo. Qualquer coisa que aumente a eficiência destas normas de conduta comunicacional está, *ipso facto*, aumentando a inteligência e fazendo isso em uma taxa de crescimento cuja magnitude não pode ser estimada de forma abstrata mas que é de maneira manifesta, às vezes tão grande, que seria difícil, se não impossível, exagerar quão grandemente a inteligência humana é aumentada através deste tipo de cooperação colaborativa.

Práticas de análise por pares autêntica e análise por pseudo-pares

O que Ginsparg conseguiu com seu sistema de arquivos e servidor automatizado foi estabelecer um espaço dialógico no qual a análise por pares ocorre em sua forma mais pura. Não quero dizer, em sua forma mais *perfeita*, que corresponderia apenas a uma possibilidade ideal e que, sem dúvida, nunca será realizada, dada a limitação humana em todos os assuntos, e uma vez que isso exigiria uma comunidade de pesquisa que fosse também perfeita em suas habilidades, e vontades, para vivenciar um conjunto de exigências extraordinariamente rigorosas, o que não se pode esperar que se concretize completamente. Mas a forma é pura em virtude da ausência em si da manipulação do discurso (e, do mesmo modo, de seus resultados no que é eventualmente aceito ou não pela comunidade de pesquisa como asserções válidas) por um mediador humano que controle *de fora* seu conteúdo ou ocorrência, filtrando-o ou moldando-o de alguma maneira não disponível para um par que seja participante do processo, em sincronia com os demais, funcionando assim em virtude de estar sob as restrições das normas comunicacionais, geralmente bem compreendidas, que governam os relacionamentos entre pares.

Agora compare isso (a análise por pares propriamente dita, a análise por pares em sua forma pura) com o que é hoje erronea-

mente considerado como análise por pares (assim como esta é assumida por seus defensores) onde somente asserções de pesquisa com o acesso controlado editorialmente deveriam ser permitidas.

(1) Quando uma análise por pares está em curso, o pesquisador faz sua asserção de pesquisa por meio de uma apelação direta a seus pares para que estes tomem ciência desta asserção, deixando esse encargo para os próprios membros da comunidade de pesquisa, que devem assumir a responsabilidade de procurar as asserções com as quais identifiquem interesse, e possam portanto ser considerados pares daquele que fez a asserção. O fornecimento do resumo (*abstract*) é o método formal pelo qual isto é feito no sistema Ginsparg. No caso da assim chamada 'análise por pares', como normalmente concebida e praticada, no entanto, o pesquisador que faz a asserção não pode efetivamente fazê-la a não ser que o editor chefe da publicação decida permitir esta ocorrência, o que é um direito não disponível aos pares do pesquisador que faz a asserção, sendo portanto uma violação subversiva do processo de análise por pares, em virtude de violar as tarefas comuns associadas ao princípio do par.

> *Comentário:* Note que isto assume que o artigo (a asserção do pesquisador) não está sendo disponibilizado a uma comunidade de pares em particular, encaminhado a ela, de uma outra maneira simultânea. Se ele está, então a ação editorial é simplesmente impertinente ao processo de análise por pares como tal, uma vez que esta ocorre, na verdade, de maneira um tanto quanto distante deste ato editorial particular, desde que o acesso seja genuinamente aberto.

(2) Se o editor do periódico faz uso escrupuloso de um revisor, ou de grupo de revisores, para decidir o que deve ser publicado, e do que deve ou não aparecer nos artigos publicados, isto ocorre por meio da seleção de um revisor que o editor julga ser um par daquele que está fazendo a asserção, sem estar certo de que quem faz a asserção está direcionando-a à pessoa que o editor escolhe para exercer a função de par. É possível que um editor escolha erroneamente como um par de quem está fazendo a asserção alguém que de fato carece de conhecimentos sobre o assunto-tema, e suas problemáticas, o que seria condição essencial para ser um par em relação àquele que faz a asserção.

Em uma análise por pares autêntica, por outro lado, à medida que ocorre uma seleção de participantes no processo de análise por pares, estes participantes são auto-seletivos, na medida em que é

possível a qualquer um engajar-se com confiança no diálogo como um par, cujas pretensões de sê-lo será honrada pelos outros participantes. Se algum indivíduo está realmente operando como um par em um processo de análise por pares autêntico, entretanto, isso é algo que só pode ser determinado empiricamente pela cuidadosa interpretação do processo dialógico, para averiguar se uma dada contribuição supositiva ao diálogo está, ou não, realmente funcionando como tal. Assim, não há na verdade nenhuma pessoa selecionando os participantes em um processo autêntico de análise por pares, sendo que a seleção é função do curso do processo dialógico em si. Em um certo sentido, a seleção é feita pela própria comunidade de pesquisa, à medida que esta participa do processo discursivo.

> *Comentário:* Imagine um caso em que alguém não qualificado para exercer um papel de par, de alguma forma tente participar do processo de diálogo crítico que se segue à divulgação de uma certa asserção. Se essa falta de qualificação não aparece a partir do que é dito pelo intruso, identificado como tal por aqueles que são pares propriamente ditos, então isto não traz nenhuma conseqüência neste contexto e é como se isto nunca tivesse realmente ocorrido. É assim em uma conversa comum, quando alguém tenta se intrometer e é simplesmente ignorado por aqueles que já estão engajados. As palavras pronunciadas falham no processo de engajamento. É claro que isto também pode acontecer com algo dito por alguém que é de fato um par qualificado, mas isto é uma das imperfeições naturais do processo dialógico. Mais importante ainda, pode acontecer também do intruso não ser identificado como tal e não ser conseqüentemente ignorado, como seria apropriado no caso de um pseudo-par, mas ao invés disso, ser tratado erroneamente com o devido respeito de um par. Uma importância e um peso indevidos podem ser dados à suposta contribuição desta pessoa. Isto pode acontecer, mas a razão pela qual a intervenção editorial é especialmente deletéria é que o prestígio do editor tende a fazer com que os pares participantes acreditem nas supostas (mas falsas) contribuições dos pares, contribuições estas que eles iriam, do contrário, simplesmente ignorar como impertinentes. Isso faz com que se acabe dando valor a uma objeção ou corroboração supositiva daquela fonte, que digamos, simplesmente não mereceria esse crédito, sendo que isso é feito num esforço equivocado de se dar o devido respeito ao editor, que acaba por interferir nas tarefas comuns do processo cujas regras do diálogo crítico do par teriam a intenção de promover.

(3) Na análise por pares editorialmente controlada, os revisores realmente utilizados no processo são, na melhor das hipóteses, não mais do que uma amostra insignificante daqueles que iriam idealmente estar envolvidos em um processo qualquer de análise por pares. Conseqüentemente, um peso totalmente desproporcional é atribuído à contribuição destes, considerando-se o número de pares realmente disponíveis em princípio para os propósitos de revisão. O raciocínio aqui é de que o editor deve usar seu julgamento na seleção dos pares revisores, que pode ser um, ou alguns poucos, mas sem um critério básico comum para fazer isto, a não ser o fato de se supor que o editor é uma pessoa de bom julgamento. Não há dúvida de que muitos editores têm de fato um bom julgamento e que sua seleção de revisores pode ser assumida como sendo razoavelmente justa. Mas uma vez que a opinião dos revisores acaba por ser determinante no processo de publicação, tendo como base somente a confiança que o editor tem neles e é, de início, o editor quem os seleciona, esconde-se o fato de que este é um sistema elitista, em que os editores, que deveriam por seu lado ser pares dos leitores de suas publicações, estão atuando na verdade como pares Orwellianos, ou seja, pares que são mais pares do que os pares a quem deveriam servir. Esses 'serviçais' são na verdade mestres na dissimulação, independente de sua qualidade como pensadores. O apoio do sistema editorial de publicação, sendo o único modo legítimo de acesso de investigadores a pesquisas em suas áreas é autoritário e radicalmente oposto ao espírito básico da pesquisa moderna, que toma o princípio do par como fundamental pelas razões explicadas acima.

(4) Nesta forma deficiente de análise por pares editorialmente comissionados, a função da comunicação é despida de toda força lógica, já que o artigo não é disponibilizado à comunidade dos pares a não ser depois que ele já tenha sido chancelado como 'correto'. Assim não há razão para o interesse na questão de como uma investigação pode ser tornada inteligente pelo desenvolvimento de normas de comunicação ou diálogo, e realmente há, de fato, pouco interesse em tais assuntos, que são considerados pertinentes somente a questões que se preocupam com a eficiência da 'distribuição' do 'resultado' da investigação. O estudo da retórica da investigação (o processo dialógico no qual a aceitação é encorajada ou desencorajada) se torna então puramente político em concepção, já que esta é pertinente somente ao uso prático do conheci-

mento já produzido, e as instituições de apoio a pesquisas são concebidas como fábricas cuja melhoria deve ser na forma de maximização da produtividade, através do uso de recompensa ou punição.

LIÇÕES A SEREM APRENDIDAS A PARTIR DO SISTEMA DE PUBLICAÇÃO GINSPARG

O que é usualmente defendido como análise por pares em publicação é na verdade uma 'paridade' Orwelliana, tanto quanto isto possa realmente funcionar como uma prática de controle crítico, onde uma classe de pessoas recebe sistematicamente um certo *status* no processo comunicacional profissional, que os posiciona como autoridades, enquanto parecem que funcionam como meros pares a serviço de seus pares. A classe privilegiada em questão não é, porém, a classe de revisores comissionados (que na verdade têm pouco poder em suas mãos) mas sim a classe dos editores chefes, que mantêm o controle das diversas mídias disponíveis, locais potenciais de publicação. São estes que comissionam revisores e definem que peso colocar, se algum, sobre suas opiniões, no processo que determina sua decisão sobre a aceitabilidade, a possibilidade de revisão e no final sobre a própria publicação das asserções de pesquisa.

Mesmo que minha abordagem aqui possa sugerir o contrário, não se coloca uma crítica geral negativa ao papel dos editores, cuja função seletiva e organizadora na pesquisa é indispensável, merecendo muito mais apreciação por seus esforços do que normalmente recebem. É que até agora, as considerações se limitaram a apontar a perspectiva equivocada com que eles são considerados, como pessoas que trabalham com o objetivo de controlar o fluxo de discurso na pesquisa, de uma posição superior à do próprio pesquisador, uma posição determinada por uma impossibilidade *de facto* em se estabelecer um meio de distribuição que pudesse prover um acesso irrestrito às asserções de pesquisa. Esta limitação não mais existe, o que foi inicialmente reconhecido, na prática, por Ginsparg, quando ele decidiu estabelecer em Los Alamos uma forma modificada do sistema até então existente de distribuição de pré-publicações. Em conseqüência, as práticas editoriais na física geral têm passado por uma rápida e continuada reforma ao longo das últimas

décadas. Há aqui a sugestão de que o problema de se estender o princípio da paridade de modo a incluir a função editorial, ao invés de se colocar um contra o outro, é abordado pragmaticamente. No momento, entretanto, as confusões sobre o controle crítico da pesquisa que foram geradas pelas concepções equivocadas sobre a análise por pares, paradigmaticamente exemplificada pela análise por pares editorialmente comissionados, fazem com que seja impossível perceber claramente o que seria uma função editorial autêntica na comunicação de pesquisa. Mas esta discussão terá que ser abordada em outro estudo.

Em todo caso, uma das coisas que podemos aprender ao refletirmos sobre a importância do sistema de publicação Ginsparg é que quando uma tradição de pesquisa alcança um estado maduro, ela não requer uma liderança editorial na vanguarda da pesquisa, na área. Reciprocamente, quando uma tradição de pesquisa não consegue fazer uso efetivo de um sistema assim, isto pode resultar da necessidade de uma liderança editorial. Neste ponto do processo de pesquisa, as exigências podem ser demasiadas para que as pessoas possam atuar efetivamente em um sistema como o Ginsparg de publicação primária, que provê um ambiente de comunicação formal com grande austeridade, livre de autoridade. É claro que deve haver outras razões para que um dado campo de pesquisa seja incapaz de fazer uso efetivo de um ambiente de publicação primária livre de autoridade como este:

> Por exemplo, pode ser que um 'colegiado invisível', controlando a circulação privada de pré-impressões de artigos, mantenha a posse exclusiva da vanguarda em um dado campo e o apelo de seus próprios interesses, como membros de uma classe privilegiada, seja simplesmente grande demais para que estes indivíduos queiram se aproveitar da oportunidade de fazer um movimento radicalmente igualitário como o que Ginsparg e seus associados fizeram em Los Alamos, ao estabelecer seu sistema de publicação automatizado e sem restrições. Havia certamente uma aposta ali, e deve ter havido um número substancial de físicos entre aqueles que inicialmente adotaram o novo sistema, e que eram inicialmente resistentes ao estabelecimento do serviço de pré-impressões de acesso aberto, acreditando que um certo idealismo pode ser bom, mas temendo que a qualidade de trabalho que apareceria, sob as condições de acesso irrestrito, poderia resultar no declínio da área. Parece razoável supor que este poderia ser realmente o caso, pelo menos para algumas áreas que poderiam provavelmente adotar de imediato o sistema

> Ginsparg com sucesso, mas às quais ainda falta uma firmeza necessária na liderança, por parte de líderes mais respeitados, para que a transição de um ambiente protegido, ao qual estão acostumados, possa ser feita para um outro ambiente, que somente pode ser 'visto com alarme'.
>
> Ou pode ser que a área seja uma daquelas na qual as fontes de financiamento são tais que as pesquisas importantes devem ser guardadas em segredo e a publicação primária deve ser censurada cuidadosamente e de maneira tática, para garantir que nada ali discutido possa arriscar o relacionamento com as principais fontes de financiamento da área, devido a violações de 'segurança' não intencionais, sejam estas fontes de caráter comercial ou governamental. O controle extensivo, e crescente, da pesquisa por interesses privados, assim como por fundos governamentais clandestinos de pesquisa para propósitos de segurança nacional é sem dúvida suficiente para explicar por que um certo número de áreas possivelmente não pode fazer uso de tal sistema e deve depender, em alto grau, de editores como censores.
>
> Ou pode ser que a área ou sub-área seja tão incipiente, ou sem foco, que um meio de publicação primária como o que o sistema Ginsparg oferece seja considerado como sendo de algum valor como lugar de publicação. Para tal área, o arquivo automatizado poderia ser somente uma coleção de artigos que podem ou não ser de interesse, mas que teria poucas razões, como coleção, para ser de real interesse, dado que não haveria ali um processo dialógico para o qual a coleção pudesse funcionar como contribuição. Neste caso haveria pouca razão, se alguma, para considerá-lo como um sistema de publicação primária.

Talvez haja outras razões, mas estas são suficientes para que não seja surpreendente que o sistema de publicação Ginsparg funcione bem somente para aquelas áreas que são suficientemente maduras, e que estão aptas a fazer um uso efetivo do sistema porque as práticas comunicacionais que já estavam governando estas áreas encontram uma boa disposição para isso.

Sugiro, então, que qualquer estudo sério de práticas comunicacionais, mesmo em áreas de sucesso como estas, irá descobrir que a comunicação que ocorre sob as bases de um sistema austero e formal como este é, de fato, uma pequena parte das práticas de comunicação de pesquisa, mesmo nestas áreas, e isso é verdadeiro independentemente das diversas áreas de pesquisa que ainda têm de se desenvolver para chegar a um nível de sofisticação que seja comparável a elas. O que torna o sistema Ginsparg incomum e interessante não é o fato de se apresentar como um paradigma

sobre o que é uma comunicação de pesquisa, mas que em virtude do sucesso de seu desenvolvimento como um sistema puro de publicação primária, somos capazes de ver, com uma claridade incomum, o que é uma análise por pares autêntica, em razão da simplicidade do sistema. Podemos então entender quais são os princípios por trás das práticas de publicação formal, e em virtude disso, podemos também ver que, se quisermos entender como funciona a comunicação de pesquisa em geral, no interesse de desenvolvê-la e aumentá-la, devemos descobrir quais tipos de práticas comunicacionais são realmente operativos no processo de investigação, nestas e em outras áreas, dado que não podemos supor, de maneira razoável, que mesmo o sucesso de áreas de pesquisa altamente desenvolvidas, como estas, possa se dever exclusivamente ou principalmente às suas práticas de publicação primária.

Embora eu ache que um entendimento apropriado do sucesso do sistema Ginsparg, que pode ser considerado uma aplicação de Inteligência Aumentada, nos revele ser este uma implementação ideal de publicação primária (ou seja, formal) computacionalmente assistida, os prognósticos mais interessantes para o desenvolvimento da Inteligência Aumentada nesta área repousam no desenvolvimento de assistência computacional para as diversas práticas comunicacionais envolvidas na atividade de pesquisa que precedem o estágio de investigação no qual a asserção formal dos achados supositivos ocorre. Um maior interesse nestes tipos de práticas comunicacionais menos formais e rigorosas tem que ser desenvolvido, entretanto, uma vez que estes devem ser entendidos em relação às práticas de publicação formal e estas têm sido tão pouco compreendidas até hoje, que não existe nenhum *framework* conceitual disponível para investigar estas e outras práticas igualmente importantes no que diz respeito a suas razões e necessidades. Minha discussão deste caso tem sido motivada pelo desejo de dar um primeiro passo nesta direção, através do desenvolvimento de alguns conceitos úteis para o entendimento do que foi, ou não foi, alcançado por Ginsparg, ao estabelecer seu sistema. A presente contribuição tem somente a intenção de sugerir o que uma contribuição mais rigorosa e completa do que esta, que espero disponibilizar em um futuro próximo, poderia ser. Avaliações críticas, com certeza, serão mais do que bem-vindas.

NOTAS

[1] A mesma distinção está implícita em um artigo anterior, embora não como uma distinção entre Inteligência Artificial e Inteligência Aumentada (ver Skagestad 1993). Se foi Skagestad o primeiro a distinguir explicitamente entre Inteligência Artificial e Inteligência Aumentada, exatamente com estes termos, tratando-os por uma distinção formal, eu não sei. A distinção, por si só, pode existir de alguma forma desde os idos de 1962 (se não antes), quando a idéia de uma inteligência computacionalmente aumentada foi descrita como 'inteligência aumentada' por Douglas Engelbart. Sua caracterização explícita como uma distinção a ser seguida de maneira genérica, utilizando um certo termo sugerido para a identificar, é mais importante do que se possa pensar, uma vez que esta estabelece uma certa estrutura formal que pode funcionar para a organização sistemática de idéias. De qualquer forma, Skagestad utiliza a distinção Inteligência Artificial/Inteligência Aumentada novamente em outro artigo, 'Peirce, Virtuality, and Semiotic', on-line no Projeto Paideia (1998): http://www.bu.edu/wcp/Papers/Cogn/CognSkag.htm

[2] Duas das áreas da Inteligência Artificial em que uma aplicação extensiva dos trabalhos de Peirce já foi empregada são a *representação do conhecimento* e a *abdução*, por exemplo.

[3] Skagestad também está ciente que esta é uma implicação posterior da afirmação, mas ele não faz uso dela quando articula sua conceitualização de Inteligência Aumentada como um paradigma.

[4] Em uma mensagem para o fórum de discussão PEIRCE-L (em 12-06-2002), Skagestad sugere que o reconhecimento explícito da distinção, utilizando os termos 'Inteligência Artificial' e 'Inteligência Aumentada', respectivamente, poderia ser atribuído ao cientista da computação Frederic Brooks, que foi citado por Howard Rheingold (1991: 37): 'Creio que o uso de sistemas computacionais para a amplificação da inteligência está muito mais poderoso hoje, e será cada vez mais poderoso no futuro, do que o uso de computadores para Inteligência Artificial [...]. Na comunidade da Inteligência Artificial, o objetivo é substituir a mente humana pelas máquinas, seus programas e suas bases de dados. Na comunidade da Inteligência Aumentada, o objetivo é construir sistemas que amplifiquem a mente humana, ao prover ajudantes computacionais que façam o que a mente tem problemas para fazer'. Observe que

Brooks fala em ‘amplificação’ e não ‘aumento’, embora o conceito pareça ser exatamente o mesmo.

[5] Links para estes artigos são encontrado no site ‘ARISBE’, sobre Peirce, <http://members.door.net/arisbe/menu/library/aboutcsp/aboutcsp.htm> Veja também a nota 1 para as URLs; veja o trabalho de Skagestad (1999).

[6] A rede semântica é a própria *world wide web*, aumentada por programas que provêm descrições processáveis dos recursos disponíveis, de forma que estas possam atuar como conteúdo referencial dos sites, e também por programas para processar estas informações de uma maneira útil aos usuários da *web*. Berners-Lee & Fischetti (2000) explicam sua visão ampla da *web*, fazendo uma avaliação informativa da maneira como a *web* se desenvolveu, tanto do ponto de vista conceitual, quanto de implementação.

[7] Notem que estou aqui me referindo à interação entre humanos e máquinas, de uma maneira distinta de outros tipos de entrada de informações originadas a partir de alguma fonte externa à maquina. Como tradicionalmente se concebe, a Inteligência Artificial deve ser distinguida por sua preocupação em desenvolver máquinas que tenham uma ‘inteligência’, no sentido de que qualquer inteligência que se manifeste, possa ser atribuída à própria máquina, sem referência ao relacionamento humano-máquina, onde a inteligência pudesse ser atribuída ao elemento humano da interação. Por exemplo, a habilidade do meu computador doméstico em me vencer no jogo de Reversi, quase toda vez que jogamos, parece ser um caso típico de Inteligência Artificial mas não de Inteligência Aumentada.

[8] Esta referência diz respeito a uma série de seis artigos entitulada ‘Illustrations of the Logic of Science’, publicados entre 1877-78 em volumes sucessivos do periódico *Popular Science Monthly*.

[9] Outra perspectiva que poderia ser assumida aqui seria considerar a lógica como uma teoria geral da natureza de uma questão.

[10] Este é um tipo especial de asserção, certamente, uma vez que ocorre dentro de um contexto de comunicação em uma comunidade de pesquisas em curso, mas ele pode trazer boas pistas sobre o entendimento do que é uma asserção, fora deste contexto especial.

[11] A razão para isso se encontra em considerações relacionadas ao conceito de ‘pares de pesquisa’.

[12] Uma tentativa de se desenvolver uma 'reconstrução racional' da aceitação em pesquisa é também fútil, da maneira explicada aqui. Isso não significa que a aceitação seja irracional, mas somente que ela não pode ser descrita algoritmicamente.

[13] Respectivamente, para marcar a origem do conceito de publicação primária, como um conceito analítico distinto, de valor inestimável para nossos propósitos, no trabalho de Joshua Lederberg, que é, tanto quanto eu o saiba, o primeiro que viu claramente o papel especial daquilo que ele chama de 'literatura primária' no curso e coreografia de uma pesquisa, utilizando tal termo para os documentos que funcionam como veículos materiais de publicação primária. Veja seu artigo 'Options for the Future', *D-lib Magazine*, May 1996: <http://www.dlib.org/dlib/may96/ 05. Lederberg.html> Lederberg não deve, evidentemente, ser considerado responsável pelas diferentes maneiras com que eu emprego aqui o seu conceito.

[14] Compare as diferenças entre prometer, não prometer e pretender prometer ou agir, como se alguém estivesse prometendo, mas subentendendo que isso seja somente uma atuação, e.g. tal qual um ator realizando uma peça teatral. As diferenças são sutis e complexas para serem discutidas aqui.

[15] O sistema de publicações está em: http://arXiv.org

Minha sugestão é que o leitor explore o site, particularmente a página: <http://arXiv.org/blurb>, onde se podem encontrar algumas estatísticas interessantes sobre o seu uso, além de diversos artigos de Ginsparg que são realmente extraordinários em tornar claro como ele e alguns de seus associados concebiam seu sistema, desde o começo. Leia o material mais antigo primeiro. Pretendo, em um estudo futuro, fornecer uma análise muito mais informativa sobre o sistema Ginsparg, como um projeto que se encontra ainda em elaboração, embora não esteja muito claro ainda o quanto foi realmente concebido pelo próprio Ginsparg.

[16] Se o ato realmente conta como uma publicação primária ou não, isso depende de se entendê-lo, ou não, como um ato de publicação primária por toda a comunidade de pesquisa relacionada, ou seja, isso depende da comunidade 'comprar' ou não a idéia. Caso a comunidade não 'compre' a idéia, dizemos que nenhuma publicação primária ocorreu realmente: a tentativa de se fazer uma asserção de pesquisa 'falhou', como diria J. L. Austin. A presunção

tem um papel certamente importante na assimilação de idéias por uma comunidade, da mesma maneira que o tem no reconhecimento de um par como tal, mas este é um tópico que não será explorado aqui.

[17] A palavra colegiado é utilizada aqui no sentido de um 'grupo de colegas'.

[18] Este processo de obscurecimento pode ser visto como um registro público nos arquivos do Fórum de Setembro, patrocinado pela revista *American Scientist*, que foi gerenciada desde o seu início por Steven Harnad, a figura mais influente em formar a opinião sobre a natureza da publicação, especialmente considerando os planos de tornar todas as publicações de pesquisa *on-line*, com acesso público, sem restrições de qualquer ordem, e gratuito. Este tem sido o principal fórum público de discussão destes assuntos, devido ao extraordinário caráter de influência de todos os que subscreviam o fórum: aqueles aspirando por mudanças (os agitadores), bem como a forte oposição ao movimento de disponibilização on-line de publicações de pesquisa. Harnad, como um cientista da computação com um alto, e merecido, prestígio em ciências cognitivas, foi o principal agente de obscurecimento da relevância do sistema Ginsparg, embora sua posição tenha sido 'oficialmente' a de um advogado de defesa do sistema (embora ele nunca tenha se autoproclamado desta maneira). Mas, como se diz: com um amigo como Harnad, quem precisa de inimigos? Ocasionalmente, tenho reportado na lista PEIRCE-L, durante os últimos anos, minha opinião sobre o que estava acontecendo ali, mas este não era o local adequado para se tentar prover um sumário do que aconteceu. Caso o leitor tenha tempo e energia necessários para navegar pelo curso da discussão, desde o tempo em que Harnad a assumiu, logo depois que esta se estabeleceu, até o presente, este a achará interessantemente informativa como uma exibição das estratégias para anular a ameaça aos modos atuais de se controlar a publicação de pesquisas, que o sistema de Ginsparg inicialmente aparentava apresentar. O leitor encontrará algumas tentativas colocadas por mim, aqui e ali, para impor alguma resistência a isso, mas a vontade de Harnad em encerrar qualquer discussão de qualquer tópico que ele considerasse uma ameaça às sensibilidades estabelecidas acabaram por tornar impossível qualquer avanço a este respeito. As seguintes URLs apresentam uma versão de todos os arquivos do fórum: http://amsci-forum.amsci.org/archives /

september98-forum.html ou http://www.cogsci.soton.ac.uk/~harnad/Hypermail/ Amsci/index.html. Espero poder trazer uma descrição detalhada do que ocorreu nestes e em outros fóruns relacionados, em outro artigo.

[19] Mas como pode alguém dizer isso? Como posso eu, em particular, não eu como um pesquisador em algum dos assuntos a que o sistema serve, fazer uma asserção como esta? Uma resposta curta seria dizer que, se a qualidade da pesquisa realmente estivesse em declínio nestes campos, em virtude de ser este um sistema de publicação primária deficiente, isso iria se manifestar na forma de uma dificuldade crescente de comunicação nestas áreas, além de uma frustração contínua e crescente que seria aparente em coisas como, por exemplo, o abandono de formas apropriadas e uma tendência ao relaxamento na preparação dos artigos para publicação, por uma falta crescente de cuidados com a sobriedade das críticas efetuadas, pela formação de facções rivais, e assim por diante, o que resultaria no abandono do uso do sistema por pesquisadores sérios, que simplesmente voltariam ao sistema anterior de distribuição de pré-impressões. Mas tais sinais de abandono e degeneração não surgiram. Ao contrário, o que ocorreu foi um crescimento constante no uso, sob uma taxa de incremento uniforme, desde o começo até hoje.

[20] Como na fábula de George Orwell sobre o autoritarismo.

[21] É verdade que as opiniões de algumas pessoas terão, na prática, inevitavelmente mais peso do que a de outras, e sem dúvida devem ter se elas estabelecem um caminho devidamente registrado que garanta isso. Mas isto deve permanecer no nível do julgamento individual e não deve ser confundido com a compreensão pública de uma dada comunidade científica, que sempre se preocupa somente com as características do assunto-tema que sejam do interesse constitutivo da comunidade de investigadores como tais. Em outras palavras, nenhuma comunidade de investigação científica como esta pode legitimamente se preocupar em classificar seus próprios membros em termos de posição e valor na comunidade porque fazer isto é perder de vista seu assunto-tema decorrendo então na introspecção do grupo.

Referências

BUSH, VANNEVAR. 1945. As We May Think. *Atlantic Monthly* 176 (1) (reimpresso em: Nyce, James M. and Kahn, Paul. 1991) <http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>

NYCE, JAMES M.; KAHN, PAUL (eds.). 1991. *From Memex to Hypertext:* Vannevar Bush and the Mind's Machine. Academic Press, Inc. pp. 641-649.

RHEINGOLD, HOWARD. 1991. *Virtual Reality*. Secker & Warburg.

SKAGESTAD, PETER. 1993. Thinking with machines: Intelligence Augmentation, Evolutionary Epistemology, and Semiotic. *The Journal of Social and Evolutionary Systems* 16 (2): 157-180. <http://members.door.net/arisbe/menu/library/aboutcsp/skagesta/thinking.htm>

__. 1996. The mind's machines: the Turing machine, the Memex, and the personal computer. *Semiotica* 111 (3/4): 217-243. <http://members.door.net/arisbe/menu/library/aboutcsp/skagesta/semiotica.htm>

__. 1999. Peirce's inkstand as an external embodiment of mind. *Transactions of the Charles S. Peirce Society*, Summer XXXV (3): 551-561.

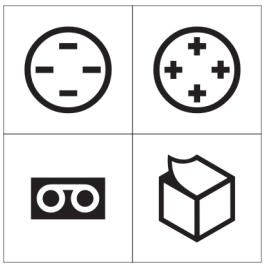
TURING, ALAN M. 1965. On computable numbers, with an application to the Entscheidungsproblem. Em: The Undecidable, Martin Davis (ed.), Raven Press. pp.116-154. Publicado originalmente nos *Proceedings of the London Mathematical Society* 1936, 2nd Series, 42: 230-265.

BERNERS-LEE, TIM; FISCHETTI, MARK. 2000. *Weaving the web:* the original design and ultimate destiny of the world wide web. Harper Collins Publishers Inc.

COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA



CAPÍTULO 2

APRENDIZAGEM *QUA* SEMIOSE

André De Tienne

INTRODUÇÃO

Como é estranha nossa condição. Como os filósofos têm demonstrado, particularmente Sócrates e Platão, não sabemos o que é a 'justiça', mas falamos dela a todo o momento; não sabemos o que significa 'ser', mas aqui está ele presente em quase tudo que escrevemos. Prosseguimos assim com cada conceito que usamos. 'Aprendizagem' não é uma exceção. Dolorosamente conscientes de nossa ignorância, precisamos 'aprender' a todo momento, desde nosso nascimento até nossa morte. O que é a 'aprendizagem'? Podem-se encontrar respostas diretas para esta questão: aumentar o conhecimento, diminuir a ignorância, adquirir uma nova habilidade, encontrar uma explicação satisfatória, compreender algum estranho fenômeno. Usamos este termo em todos estes sentidos, e não há nada de difícil em capturar seu significado. Aprender é parte de nossa experiência humana, e estamos todos bem familiarizados com esta atividade. 'Aprendizagem' é apenas mais uma dessas pa-

lavras que utilizamos para caracterizar, de maneira expediente, uma dimensão permanente de nossa vida, sem muita precisão. Mas, como ela se aplica a situações variadas, poderíamos naturalmente sugerir que, em algum lugar, deve haver um fundamento comum que, por mais vago que possa ser, demande uma cautelosa investigação e análise.

Em seu uso mais freqüente, a aprendizagem está vinculada à aquisição de conhecimento e, portanto, a uma apreensão da realidade que luta por se tornar cada vez mais verdadeira. Pode-se considerar que aprender, desta maneira, está conectado necessariamente à noção de verdade, mas 'verdade' aqui não deve ser considerada no sentido latino de *veritas*, mas no sentido grego de *alêtheia*, como Heidegger insistia, ou seja, como um processo de desocultamento. Para Platão, o dado da experiência ordinária é um véu que precisa ser removido, e neste processo nos defrontamos com a apreensão intuitiva das formas ideais, que gravitam no mundo do ser, bem além deste nosso decepcionante mundo de mudanças e do vir a ser. O conhecimento acaba sendo o *epistêmê* ou a *noêsis* do estável, imaculado, abstração pura; idéias completamente reveladas, trazidas à luz da *alêtheia*, por si só uma emanação do bem último. Mas este conhecimento intuitivo é um privilégio de algumas poucas almas, altamente treinadas no exercício filosófico. Nós, seres humanos ordinários, estamos condenados a viver acorrentados no fundo da caverna, convencidos de que o mundo não se estende além do fenômeno das sombras, que constituem nossa percepção. 'Educação', diz Platão por meio de Sócrates, 'não é o que algumas pessoas declaram ser, dar conhecimento a almas que não o têm, como se déssemos visão a um olho cego ... O poder para aprender está presente nas almas de todos e o instrumento com o qual aprendemos é como um olho que não pode se orientar da escuridão para a luz sem que movimentemos todo nosso corpo ... A educação toma, como garantido, que a visão está na alma, mas que essa não está orientada na direção correta, ou seja, olhando para o que deveria olhar, e tenta redirecioná-la apropriadamente' (*República VII*, 518c, d). Ignorância, ou *agnoia*, é, para Platão, o poder de olhar para a direção errada. Aprendizagem é o processo por meio do qual nos tornamos atentos deste erro, e damos os passos para remediá-lo. O corpo todo então precisa girar. Girar a cabeça, enquanto permanecemos sentados acorrentados ao fundo da caverna, não é o suficiente.

Aprendizagem implica em constante desaprendizagem. Outros filósofos, além de Platão, expressaram essa mesma idéia. Entre eles, Charles S. Peirce, para quem a aprendizagem envolve um movimento permanente de fuga das quatro barreiras que bloqueiam o caminho da investigação: a) fazer asserções que vão além do que realmente sabemos, b) afirmar que há fatos que não podem ser conhecidos, c) afirmar que há fatos que não podem ser explicados e c) afirmar a infalibilidade (EP2: 49-50). Peirce pode não ser um seguidor de Platão, mas certamente encontraremos, aqui e ali, em seus escritos, traços de grande simpatia pelo idealismo do fundador da Academia. Platão cometeu dois erros, segundo Peirce: o primeiro foi ver o principal valor da filosofia em sua influência moral, e o segundo foi assumir que o objetivo último da vida humana seria a aquisição das idéias puras. Entretanto, estes dois erros se equilibram tão bem que, tomados em conjunto, 'acabam por expressar uma visão correta dos propósitos últimos da filosofia e da ciência em geral' (EP2: 38). A conferência de Peirce, em 1898, 'Philosophy and the conduct of life', termina com as seguintes palavras: 'As partes mais profundas da alma somente podem ser atingidas através de sua superfície. Desta maneira, as formas eternas, com as quais a matemática e a filosofia, e também outras ciências, irão, por um lento processo de filtragem, gradualmente penetrando em nosso ser, e assim irão influenciar nossas vidas, e assim o farão ... porque são verdades eternas e ideais' (EP2: 41). Essas partes mais profundas da alma são domínios do sentimento e do instinto, as fontes de nossas motivações, e a real inspiração para a direção que escolhemos para dar a nossas vidas. Nosso instinto é muito menos falível do que nossa razão superficial e é, tanto quanto esta, capaz de desenvolvimento e crescimento, por meio da experiência, especialmente aquela parte da experiência que é filtrada através da razão cognitiva. A idéia Peirceana de 'aprendizagem' faz eco em Platão, ainda que abafado. Peirce entendia que o progresso das ciências ocorria na medida em que ficavam cada vez mais abstratas, em suas matematizações. A finalidade da matemática é descobrir o mundo real potencial, o cosmos do qual nosso mundo é apenas um locus arbitrário (EP2: 40). O mundo potencial real é o domínio das idéias de Platão, com uma diferença essencial: é um mundo que incorpora a continuidade. As verdades ideais e eternas não são desconexas, nem discretas, e são vivas – elas crescem e evoluem. Como o 'bem' de Platão, elas podem se transformar em outras idéias, mas ao contrário das idéias de Platão, elas precisam

de tempo para amadurecer, e seu destino se sujeita ao capricho do acaso. A insistência de Platão sobre o mundo das idéias como o único legítimo pretendente ao título do 'ser' levou-o a desconsiderar o mundo em que vivemos.

Peirce recusa esta tentação por não estar preso à ilusão platônica de que seria possível a alguns humanos atingir a *noêsis* das idéias, equivalente a intuição pura das formas em si mesmas. Nenhuma intuição, nenhuma redução eidética *à la* Husserl, é possível com Peirce. As verdades eternas são reais, independentes do que pensamos que sejam, e todo o processo de aprendizagem consiste em nos aproximarmos cada vez mais desta realidade. Mas os meios para alcançá-la estão além da compreensão de Platão, e seus seguidores. 'Uma idéia pura sem uma metáfora ou outra vestimenta significativa é uma cebola sem pele' (EP2: 392). Para Peirce, idéias não podem existir sem serem consubstanciadas, caso contrário evaporam no ar. Uma substanciação é essencial, mas toda a arte é fazermos com que esta seja tão translúcida quanto possível. Voltemonos agora a esta questão.

Em um artigo apropriadamente *intitulado* 'Toward a Peircean semiotic theory of learning', Nathan Houser (1987) expressa sua crença de que a teoria dos signos de Peirce é 'de fundamental importância para uma teoria correta da aprendizagem', concordando com Charles Morris sobre o fato de que o que dá à semiótica Peirceana um poder especial de explanação é seu 'foco na estrutura triádica da ação sígnica', sendo uma das conseqüências a capacidade de como uma teoria completa, consistente e singular é capaz de explicar fatos óbvios relacionados à aprendizagem, tais como o papel desempenhado pelo conhecimento de fundo, ou o papel das metáforas e das analogias (Houser 1987: 270-71). Essas são asserções poderosas. Que a triadicidade dá à teoria de Peirce um poder especial, não há necessidade de maiores esforços de defesa. Isto já foi suficientemente demonstrado, mesmo matematicamente. Ao contrário de seus contemporâneos, Peirce foi um lógico que entendeu profundamente a proeminência ontológica das estruturas lógicas.

O primeiro artigo importante que Peirce publicou, 'On a new list of categories' (CP 1.545, EP1: 1-10), em 1867, foi o resultado de dez anos de árdua pesquisa, no qual ele firmemente estabeleceu a estrutura universal da representação em geral. Esta estrutura foi descrita como irredutivelmente triádica. Ela envolvia, primeiro, o isolamento de um elemento que incorporava o fundamento da re-

presentação — um elemento que carregava, em si-mesmo, o poder de estar para alguma outra coisa de modo a trazer de volta a sua presença (o *quale*, seja este uma relação monádica, diádica ou triádica); segundo, outro elemento que já tinha sido representado pelo *quale*-signo anterior, antes de sua realização atual (o *correlato*); e terceiro, um elemento cuja tarefa principal seria reconhecer que a realização corrente pertence à mesma classe da realização passada, assim como encontra-se representada pelo correlato (o *interpretante*). Uma característica crucial desta análise era a aceitação de que nenhuma representação poderia ter lugar no vazio, ou seja, que toda representação sempre emergeria em um *continuum*, que não poderia ser abstraído de sua definição. O principal motor deste *continuum* reside no interpretante, um elemento lógico cujo principal papel é o de ser um mediador[1] de comparação e reconhecimento.

Embora Peirce, em definições maduras do conceito de signo, se desvie dos principais conceitos apresentados na definição anterior (a referência a um correlato, por exemplo, foi integrada por um refinamento da noção de interpretante, e substituída pelo objeto), o papel central atribuído ao interpretante nunca foi negado. Uma terceira característica essencial é a total ausência de psicologismos na análise. O fato da lógica anteceder a psicologia é um dos princípios fundamentais da filosofia de Peirce, o que as vezes é difícil de ser compreendido pelos psicólogos, mesmo hoje, especialmente considerando que a psicologia de hoje não é a ciência que Peirce conheceu na virada do século XX. Mas é importante entender este fato claramente. A estrutura representacional é independente do conceito de mente. Quando Peirce fez a descoberta fundamental de que todos os pensamentos eram signos, ficou claro que a 'autora' da representação não era uma mente, mas que as representações, ao acontecerem, acabavam por constituir uma mente. Signos são a condição de possibilidade do fenômeno mental. Para compreender a vida da mente é necessário primeiro entender a vida dos signos. Esta não era simplesmente uma metáfora para Peirce. Não podemos nos esquecer que signos não são entidades ou substâncias discretas e inertes, mas estruturas dinâmicas relacionais; tendemos a não vê-las como 'vivas' — como parte da fábrica do *continuum* — por causa das lentes deformadoras de nossa análise abstrata. Este é precisamente o objeto da Lógica (ou semiótica, como Peirce a chamava). É apenas porque nossa única experiência do que é mental está confinada à nossa própria mente, ou, mais

amplamente, à mente social na qual participamos, que é difícil, para nós, imaginar que poderia haver um processo do tipo mental que não tem lugar dentro de uma 'mente', como a concebemos. Mas Peirce, em diversos momentos, apesar de se recusar a 'atirar um osso para Cerberus' (em suas próprias palavras, quando se resignou ao falar de 'intérprete' no lugar de 'interpretante', de modo a ser parcialmente compreendido), prefere utilizar o termo 'quasimente', um termo técnico usado expressamente para indicar que o termo mais familiar 'mente' é apenas uma instanciação especial de um fenômeno mais geral, e que a lógica (ou semiótica) se incumbe de analisar não somente o funcionamento da mente humana. É essencialmente por esta razão que a semiótica deve preceder à psicologia, seja quando nos reportamos a uma Psicologia individual tradicional ou a uma psicologia 'social'. Esta última é mais semioticamente atenta que a primeira, mas isto não muda o fato de que é ainda focada numa instanciação especial, uma instanciação social do conceito mais Peirceano, mais genérico de 'quasi-mente'.

Esta é uma revolução Copernicana de Peirce: aquilo que experimentamos como 'mente' (seja social ou não) não é o que é porque se serve de signos, mas porque é feito de signos. Ser mental é ser totalmente permeado de signos. Quando esta vida ganha um padrão distinto, podemos então chamá-la, por exemplo, de humana, como em oposição à outra coisa, tal como, por exemplo, uma mente símia. A semiótica Peirceana é mais um estudo da 'quasimente' do que de instanciações acidentais, por mais tentadora que seja uma instanciação. Não quer dizer que Peirce não fale da mente humana. Ele o faz a todo o momento mas sempre de uma perspectiva mais abrangente.

Que nenhuma teoria da aprendizagem poderia dispensar a semiótica é uma evidência para Peirce. Na seqüência deste capítulo, iremos demonstrar porque este é o caso. Façamo-lo examinando o que Peirce tem a dizer sobre a natureza da aprendizagem, numa referência especialmente reveladora em 'On Topical Geometry, in General' (CP 7.536, c. 1899).

> Todo fluxo de tempo envolve aprendizagem; e toda aprendizagem envolve o fluxo de tempo. Assim, nenhum *continuum* pode ser apreendido, exceto por meio de uma geração mental dele, de uma idéia de algo que se move através dele, ou de algum modo equivalente a isso, e fundamentado nisso. [...] Assim, qualquer apreensão da idéia de continuidade envolve a consciência de aprendizagem. Em seguida, toda apren-

> dizagem é raciocínio virtual; [...] Para nos convencermos de que toda aprendizagem é raciocínio virtual, temos apenas de considerar que a mera experiência de sentir-reagir não é uma aprendizagem. Esta é apenas uma experiência a partir da qual alguma coisa pode ser aprendida, desde que interpretada. A interpretação é a aprendizagem. Caso se tente objetar afirmando que deve haver algo como uma primeira coisa aprendida, eu replico que isto é como dizer que deve haver uma primeira fração racional, na ordem das magnitudes maiores do que zero. Não existe um tempo mínimo que uma experiência de aprendizagem deve ocupar. Pelo menos, não concebemos assim, quando concebemos que o tempo é contínuo; para cada fluxo de tempo, por mais breve que seja, há uma experiência de aprendizagem [...]. Assim, cada raciocínio envolve outro raciocínio, que por sua vez envolve outro, e assim até o infinito. Cada raciocínio conecta alguma coisa que acabou de ser aprendida com conhecimentos já adquiridos, de forma que assim aprendemos o que não sabemos. [...] Raciocinar é uma experiência nova que envolve algo antigo e algo ainda desconhecido. O passado, como aqui colocado, é o ego. Meu passado mais recente é o meu ego predominante; meu passado distante é meu ego mais generalizado. O passado da comunidade é nosso ego. Quando atribuímos um fluxo de tempo a eventos desconhecidos, imputamos um quase-ego ao universo. O presente é a representação imediata que estamos justamente aprendendo, e que nos traz o futuro, ou o não-ego, de forma a ser assimilado ao ego. Podemos então ver que a aprendizagem, ou representação, corresponde à terceira categoria Kaino-pitagórica.

Peirce está aqui sustentando um discurso que é, ao mesmo tempo, lógico e metafísico, e portanto pré-psicológico.[2] Cinco das asserções de Peirce devem ser aqui melhor examinadas: (1) que existe uma relação essencial entre aprendizagem e o fluxo do tempo; (2) que a aprendizagem é um processo contínuo; (3) que aprendizagem é raciocínio virtual; (4) que aprendizagem é interpretação; (5) que aprendizagem é representação, e portanto outro nome para Terceiridade, a terceira das categorias de Peirce.

APRENDIZAGEM E TEMPO

Que a aprendizagem de qualquer coisa consome tempo é uma asserção trivial. Mas existe aqui uma idéia muito menos trivial. A aprendizagem faz parte da fábrica do tempo. Como assim? Incluídas na idéia de aprendizagem estão as idéias de crescimento e desenvolvimento (poderíamos utilizar aqui o adjetivo 'mental', mas isso é de

menor importância) e, portanto, a idéia de processo. A consciência de um processo é o que eminentemente caracteriza a cognição (CP 1.381). Como Menno Hulswit já observou, um processo, para Peirce, é 'uma seqüência contínua de eventos, que deriva sua unidade, ou ordem interna (o que a distingue de outros processos), de uma causa final, que direciona a seqüência para um estado final, que por sua vez pode evoluir' (Hulswit 1998: 195). Cada 'evento' no processo é um 'momento' particular dentro dele, no mínimo um infinitesimal segmento deele, mas que contém elementos relacionais suficientes de forma que possa ser identificado como uma parte da história dinâmica interna do processo, sendo uma parte que, como um todo, mostra uma consistência interna suficiente para ser suscetível de abstração e representação.

Um evento não pode ser adequadamente isolado de eventos anteriores e posteriores sem perder suas características essenciais, uma característica de 'emergir a partir de', ou 'levando a', que são responsáveis pela continuidade do processo. Assim, um evento não é o resultado de uma abstração, retirado do fluxo do tempo, mas um elemento dinâmico constitutivo do fluxo. Peirce faz uma distinção bem clara entre um evento e um fato, sendo um fato o que pode ser abstraído de um intervalo de tempo e representado por meio de uma proposição, pelo poder do pensamento. Fatos são representações discretas, eventos não são. Um processo é uma seqüência contínua de eventos, ganhando assim uma identidade peculiar (sua ordem interna), que Hulswit chama de causa final. Uma das maiores contribuições de Hulswit é exatamente insistir na manutenção do conceito Peirceano de causa final. Causas finais, como ele mostra, não são eventos futuros causando eventos presentes, mas possibilidades gerais que podem se concretizar no futuro. Desta forma, são leis gerais que ditam as direções gerais que seqüências particulares de eventos devem seguir, de tal forma que o processo constituído por esses eventos possa ganhar uma identidade crescente à medida que o tempo passa, sendo essa identidade uma corporificação da idéia geral representada pela causa final.

Como Peirce explica em 'The law of mind' (EP1: 331), nenhuma idéia geral pode ser apreendida em um instante, mas deve ser vivida no tempo; ela permeia cada intervalo de tempo com a sua presença viva. Uma idéia geral determina eventos numa perspectiva que não é completamente previsível. A referência ao futuro é um elemento essencial de qualquer processo. Como Peirce afirma,

caso o fim de um processo já esteja explicitado, não há espaço para desenvolvimento, para o crescimento, para a vida. Uma causa final apenas indica uma tendência definida, mas não tem o poder de ditar a concatenação precisa de ações e reações, de forma que o futuro venha como esperado. Hulswit nos diz que a causação final tem dois sintomas: (a) o estado final de um processo pode ser atingido de diferentes maneiras, e (b) o processo é irreversível (Hulswit 1998: 79, 94).

Se alguém decide cozinhar uma torta de maçãs, a idéia geral de uma deliciosa torta de maçãs irá guiar uma seqüência de ações que tenderá a produzi-la, mas não ditará precisamente qual receita usar, qual a quantidade de quais ingredientes devem ser misturados, e em que seqüência, tempo de cozimento, etc. Todos esses fatores podem variar (dentro dos limites permitidos pela idéia geral) mas o resultado final, seja com sabor de canela ou não, ainda constituirá uma torta de maçãs, ou seja, um resultado que pertence ao tipo geral representado na causa final. E uma vez que a torta está cozida, não há como reverter o processo e destilar dele os ingredientes originais. O mesmo acontece com a aprendizagem, se concordarmos com Peirce que a aprendizagem é uma propriedade fundamental de qualquer coisa que cresce no tempo. A essência da aprendizagem consiste tanto na apreensão da tendência geral que sugere uma direção para o futuro quanto na implementação criativa, ou atualização desta sugestão percebida. (Aqui, começamos a compreender em que sentido Peirce se referia à filtragem de verdades ideais e eternas: a natureza desta filtragem tem muito a ver com a noção de causação final). Se este é o caso, então a aprendizagem torna-se uma característica do universo, caso aceitemos a visão de Peirce de que as leis da natureza são produto da evolução e estão sujeitas ao crescimento. A natureza, como um todo, corresponde a uma implementação contínua aleatória de regras condicionais gerais que determinam as formas possíveis que são oferecidas para atualização. A aprendizagem cresce dentro dos limites de um plano geral condicional.

APRENDIZAGEM E CONTINUIDADE

A aprendizagem é um processo contínuo. Dada a definição anterior de processo, isso é evidente. Mas Peirce afirma mais do que isso. A aprendizagem consiste na apreensão de um *continuum*, e

todo *continuum* é uma idéia geral — 'Continuidade e generalidade são a mesma coisa [...] Tempo e espaço são contínuos porque incorporam condições de possibilidade e o possível é geral; e continuidade e generalidade são dois nomes para a mesma ausência de distinção entre individuais' (CP 4.172). A experiência, quando não refletimos sobre ela, nem em um resíduo de análise, como um nome que damos àquilo que constitui a permanente textura da vida, é um *continuum*. Qualquer experiência incorpora condições de possibilidade e oferece, constantemente, renovado vigor ao fluxo fenomenal que chamamos de presente.

Uma condição de possibilidade é uma lei que tem uma estrutura condicional formal: se uma certa seqüência de eventos ou processo tem lugar em uma ordem pertencente à alguma classe definida de ordens, então o processo terminaria tendo uma certa característica definida. Qualquer fato particular (um fato é um aspecto de um evento que foi abstraído do fluxo e colocado em uma forma proposicional) parcialmente preenche (materializa) uma predição condicional. Por exemplo, supor que algum objeto, no escuro, é vermelho, é supor que, se fosse iluminado, sua superfície iria absorver todos os comprimentos de onda de luz exceto aqueles pertencentes à porção vermelha do espectro — 'A mais insignificante das idéias gerais sempre envolve predições condicionais ou requer, para seu atendimento, que eventos ocorram, e tudo que ocorrer deve atender completamente seus requisitos' (CP 1.615). Uma predição condicional expressa uma lei, uma certa ordem geral de coisas, um hábito. Essas leis são reais, no sentido de que causam efeitos. Entretanto, elas não são causas eficientes, uma vez que não têm o poder de fazer as coisas acontecerem. Mas são causas finais, como vimos anteriormente.

Ora, Peirce afirma que 'uma vez que as idéias venham em conjunto, tendem a se fundir formando idéias gerais; e uma vez que estejam geralmente conectadas, idéias gerais governam essa conexão; e estas idéias gerais são sentimentos vivos que emergem' (EP1: 327). O poder da generalidade reside nas conexões que esta governa, e conexões são a tecitura dos *continua*. Aprendizagem é a apreensão das leis que governam as conexões. Um 'sentimento vivo que emerge' é a atenção que acompanha a conectividade crescente entre as idéias, signo de que esta conectividade não é randômica, uma coincidência, mas uma associação que obedece um princípio télico mais alto.[3] Essa atenção é viva no sentido, não

somente de seu crescimento, mas também porque constantemente adapta a tradução (atualização) do princípio télico às suas circunstâncias existenciais, provocadoras de erros, circunstâncias mutacionais. Seja o que for que ocorra, falhará levemente para preencher os requisitos, mas nunca completamente e, provavelmente, cada vez menos; e isso, considerando a possibilidade de um universo atual, é bom o suficiente.

APRENDIZAGEM E VIRTUALIDADE

'Aprender é raciocinar virtualmente.'[4] Todo raciocínio é aprendizagem, e qualquer coisa que tenha a estrutura de um raciocínio, sem que notemos isso, pelo fato de que está 'muito tênue na consciência', sem poder portanto ser criticado ou corrigido, também é aprendizagem. Por que? Porque o raciocínio é a passagem de uma crença para outra. Qualquer raciocínio, seja abdutivo, dedutivo ou indutivo, é composto por uma seqüência de proposições (premissas), por meio das quais alguma idéia que, ou não é ainda conhecida ou, por uma gradação qualquer, não foi ainda totalmente revelada, é trazida à luz em virtude de uma seqüência de premissas. Cada premissa representa uma crença de algum tipo, particular ou universal, e a representa não somente de maneira isolada, mas como uma asserção que ocupa uma posição muito bem identificada em uma ordem maior. Uma premissa é uma crença que clama por outra em virtude de sua própria associação com outras crenças já posicionadas em uma seqüência cuja identidade geral é ditada pelo que Peirce algumas vezes chama de 'princípio guia'. O princípio guia é o hábito do pensamento que determina a passagem de uma premissa para uma conclusão (CP 3.160). Existem diferentes tipos de hábitos do pensamento, e Peirce distribui-os entre os três tipos principais de inferência: abdução, dedução e indução.

Peirce chama de 'coligação' a mistura de premissas que ocorre anteriormente à conclusão, seguindo Whewell — 'A coligação é uma parte muito importante do raciocínio, chamando-nos à genialidade talvez mais do que qualquer outra parte do processo' (CP 2.442). Isso ocorre pois, uma vez que as premissas tenham sido coligadas, formando uma proposição composta, a conclusão segue de maneira quase automática, obedecendo ao princípio guia. Assim, a arte do raciocínio reside menos no fato de atingirmos a conclusão, do que na mistura que fazemos das premissas: uma coligação será somente tão efetiva quanto o princípio que a rege, em primeiro

lugar. Para serem efetivas, idéias que são coligadas não devem somente co-existir, mas devem copular de forma a conceber um descendente, razão pela qual Peirce costuma chamar a proposição composta formada pelas premissas de 'proposição copulativa'. Uma copulação gratuita poderá, ou não, ser fértil, mas uma copulação sob controle tem uma chance muito maior de atingir uma conclusão. Assim, é importante que o coligador dê à associação de crenças uma certa forma, uma forma inspirada a partir daqueles hábitos do pensamento que são inferenciais, pois estes são os mais prováveis de gerar um novo pensamento, uma nova crença: são aqueles capazes de nos conduzir do conhecido ao desconhecido. A aprendizagem, sob este aspecto, está fortemente conectada à arte de prestarmos atenção aos princípios gerais e deixá-los atuar, para que filtrem o raciocínio.

Ora, como já inicialmente explorado, raciocinar é por si só um processo e é contínuo. Diversas vezes, Peirce insiste na importância deste fato. É verdade que não se pode sustentar que toda cadeia de pensamentos seja puramente inferencial. Mas qualquer inferência, como tal, em sua própria natureza, exibe continuidade interna, uma vez que sua conectividade é governada por um princípio geral. Adicionalmente, as premissas coligadas em uma proposição copulativa têm, por si só, uma história. Como premissas, devem ter ganho seu crédito em seu passado representacional, o que significa que elas, por si próprias, foram em algum momento conclusões de outras inferências, mesmo que somente perceptuais (i.e., abdutivas, não passíveis de crítica). Nenhuma inferência está puramente isolada: podemos dizer, em certo sentido, que uma inferência é um tipo de evento, como definido anteriormente, constituindo uma porção infinitesimal do processo conhecido como raciocínio. O raciocínio como um todo é, de algum modo, um *continuum*, embora mais complexo do que as inferências lógicas. Peirce escreve:

> Não há nenhuma necessidade para supormos que o processo do pensamento, como acontece na mente, esteja sempre segmentado em argumentos distintos. Um homem segue em seu processo de pensamento. Quem é que seria capaz de dizer qual é a natureza desse processo? Ele não pode, uma vez que, durante o processo, esteve ocupado com o objeto sobre o qual estava pensando, não consigo, nem com seus movimentos. [...] De maneira prática, quando este homem se dispuser a estabelecer como teria sido esse processo, depois que o processo tenha sido concluído, sua primeira atitude será perguntar-se a que conclusão chegou. Este resul-

> tado, que ele formula em uma asserção, nós assumimos, tem uma espécie de similaridade com a atitude de seu pensamento no cessar do movimento. Tendo verificado isto, ele a seguir pergunta a si-mesmo como pode estar tão certo sobre isto; e ele procede então a procurar uma sentença que possa ser expressa em palavras e que o afetará de modo semelhante a alguma atitude prévia de seu pensamento e que, ao mesmo tempo, estará logicamente relacionada à sentença que representa sua conclusão, de tal modo que se a premissa-proposição for verdadeira, a conclusão-proposição seria, necessariamente ou naturalmente, verdadeira. [...] Mas o auto-observador não tem qualquer garantia de que esta premissa representou uma atitude na qual o pensamento tenha permanecido disponível, mesmo por um instante. [...] Adotando esta idéia, o argumento lógico somente representa a última parte do pensamento, porque supõe uma premissa que representa alguma atitude do pensamento que somente possa ter sido resultado do ato de pensar. Agora, se você separa a última parte de um tempo, você deixa um tempo anterior. Se você separa a última parte deste, ainda deixa um tempo anterior, e não há qualquer possibilidade de separação em tantas partes finais do que resta, de forma que a partir do que sobra, nenhuma parte final possa ser separada. Conseqüentemente, não há necessidade de uma série de argumentos que represente um curso de pensamento para ter um primeiro argumento, antes dos quais não havia qualquer argumento no pensamento, no sentido de que não havia qualquer argumento, afinal, no processo de pensar. Assim, nada nos impede de supor que o processo de pensamento era um processo contínuo (apesar de indubitavelmente variado) (CP 2.27, 1902).

Todos os tipos de idéias são concebíveis na cadeia do pensamento, não importando se legitimamente (inferencialmente) ou ilegitimamente. É impossível reconstruirmos, de maneira confiável, depois que uma dada cadeia do pensamento tenha seguido seu curso, todas as suas partes sucessivas, a menos que através de um meio simplificado de abstração. A abstração permitirá ao 'auto-observador' repensar esquematicamente a última parte de sua jornada de pensamento, distinguindo a conclusão das premissas. Mas isso se dará ao custo de quebrar a continuidade original; a continuidade inferencial será preservada talvez, mas apenas como um ícone empobrecido do processo original. O que Peirce afirma é que raciocinar, como um todo, é parte do denso *continuum* da experiência e, como tal, tem uma vitalidade e uma riqueza que vai além do que poderíamos capturar em palavras. A pobreza das palavras força o auto-observador a simplificar esta realidade, podendo levá-lo a con-

cluir, ilusoriamente, que seqüências de argumentos distintos (ou discretos) constituem, certamente, a fábrica do raciocínio. Tal raciocínio, então, produz conjecturas sobre a natureza do que poderia ser o ponto de partida da cadeia do pensamento, o argumento inicial que originou todo o resto, a primeira premissa. Mas não necessariamente. Como é infrutífero falar do ponto inicial do tempo, também o é falar do ponto inicial do processo de pensamento. A descontinuidade de abstrações pode permití-lo, mas apenas porque esquecemos que se trata de uma abstração. Aqui, começamos a observar que outra dimensão da aprendizagem, como uma propriedade *quasi-mental*, pode também ter a ver com o fato de que nos tornarmos atentos da real natureza da passagem do *continuum* de uma experiência vivida para o *continuum* empobrecido de representações (ou signos), que luta para reproduzir sua mais rica fonte fanerônica (*phaneron*).

APRENDIZAGEM E INTERPRETAÇÃO

'Aprendizagem é interpretação.' Isso indica que tipo de operação a filtragem de verdades significa. Foi estabelecido que isso é, em parte, uma questão de raciocínio e coligação de premissas. Coligar premissas é arranjar proposições de modo que elas se tornem um todo dotado de poder copulativo; este poder não é nada mais que o poder de um signo para determinar um interpretante. Peirce explica em vários lugares que a conclusão de um argumento é o interpretante de sua premissa. Foi em 1866 que ele percebeu isto pela primeira vez:

> Um interpretante é alguma coisa que representa uma representação a representar aquilo que ela própria representa. Aquilo que, então, apela ao interpretante - ou seja, é construído intencionalmente de forma a desenvolver uma redeclaração por parte de um outro, ou um consentimento - é um argumento, um silogismo *minus* a conclusão, posto que a conclusão de um silogismo não é parte do argumento, mas concorda com este, o interpretante (W 1:478).[5]

A concatenação proposicional que forma a premissa se tornou um signo unificado e, desta forma, clama por uma nova representação que é chamada de 'equivalente' nos primeiros escritos de Peirce, e que, quando vem, marca a premissa com o selo do reconhecimento. Esta habilidade de convocar algo, de requerer algo, é o que dá força a um símbolo. O interpretante-Conclusão recoloca a premissa

coligada de uma nova forma e lhe imputa um aumento de informação 'supérfluo' (supérfluo por não tender a aumentar nem a extensão nem a intensão do que está contido na premissa) confirmando, assim, sua significação. Peirce fala de uma aprovação, ou seja, do consentimento do interpretante em responder ao apelo do argumento. O interpretante não está satisfeito em meramente repetir a premissa de uma forma contraída; o 're-estabelecimento' inclui também a afirmação de que a representação feita pela premissa é similar àquela da conclusão.

A missão específica da conclusão é afirmar sua equivalência com a premissa coligada. Mas tal afirmação não pode ser feita sem chamar a atenção para o princípio-guia, que governa e dá identidade à inferência. Certamente, uma conclusão não é uma proposição isolada. Ser uma conclusão confere um status especial a uma proposição, um status que não é imanente a esta, mas transcendente, e que é paralelo ao status que um signo adquire ao tornar-se um interpretante. O que é este status? Encontramos seus primeiros ecos no estudo feito por Peirce em 1857 sobre Friedrich Schiller, e que o levou a distinguir três 'proto-categorias', como podemos chamá-las, aquelas do 'I' (eu), 'It' (o outro) e 'Thou' (o outro, no sentido respeitoso). Neste estudo, depois de conectar o 'I' ao Intelecto e ao princípio masculino, o 'It' à Sensibilidade e ao princípio feminino, e o 'Thou' ao Coração e também ao amor, Peirce descreve em uma nota de rodapé (W 1:15 n. 3) o 'resultado notável' de que o coração não é a mera conjunção do intelecto e da sensibilidade, mas o resultado necessário de sua união, assim como em aritmética o 7 é o resultado da soma de 3 e 4, sem estar reduzido à sua mera adição. Assim, o terceiro elemento não é simplesmente a mistura de dois elementos 'paternos', mas o resultado necessário que contém um elemento adicional não redutível às suas conjunções.

A união das premissas, não importando se a chamamos de copulação ou de coligação, deve produzir uma descendência que é a conclusão, e esta descendência não pode ser reduzida simplesmente às premissas: uma vez gerada no *continuum*, ela adquire uma alma própria, sendo dotada de seu próprio poder de crescimento. Mas esta nova alma, uma vez que descende da união de outras almas, é geneticamente marcada por elas. Algum elemento tem sido transmitido a ela de acordo com um princípio genealógico. O interpretante é o que é, possui o status que possui, em virtude deste elemento

genealógico. O fato de ser 'equivalente' não o torna 'idêntico', mas passível de exercer, por sua própria autoridade adquirida, a transmissão dos elementos que foram passados a partir dos signos determinantes. Pode-se chegar ao 7 adicionando-se o 3 ao 4, mas o 7, como os Pitagóricos bem o sabem, apesar de sua herança, tem uma vida própria. Quando uma conclusão contrai as premissas em si mesmo (com a eliminação de termos intermediários), ela se torna um novo ser, com um passado e um futuro.

Se 'aprendizagem é interpretação', isto implica na arte de obter novos interpretantes e de cuidar deles (assim podem continuar o trabalho de transmissão). Para isto é necessário que qualquer coisa que se aprenda seja compreendida como um signo ou, pelo menos, se comporte como se soubesse que é um signo — 'A palavra ou signo que o homem usa é o homem, ele mesmo. Pois ... o fato de que todo pensamento é um signo é, tomado juntamente com o fato de que a vida é uma cadeia de pensamento, prova de que o homem é um signo' (EP1: 54). Mas o que exatamente está sendo transmitido das premissas à conclusão, do signo para o interpretante? O que a interpretação significa? Aqui nós devemos nos voltar para um pequeno e conhecido texto que é bastante sugestivo:

> Para o propósito desta investigação, um signo pode ser definido como um *medium* para a comunicação de uma forma. [...] Como um *medium,* o signo está essencialmente em uma relação triádica, com o objeto que o determina e com o interpretante que ele determina. [...] O que é comunicado do objeto através do signo para o interpretante é uma forma. Não é uma coisa singular; porque se uma coisa singular estivesse primeiro no objeto e, posteriormente, no interpretante, fora do objeto, teria então de deixar de estar no objeto. A forma que é comunicada não deixa, necessariamente, de estar em uma coisa quando vem a estar em outra diferente, porque o seu ser é um ser do predicado. O ser de uma forma consiste na verdade de uma proposição condicional. Algo seria verdade sob certas circunstâncias. A forma está no objeto, onticamente, nós podemos dizer, significando aquela relação condicional ou — seguindo do conseqüente sobre a razão — que constitui a Forma e é literalmente verdade do objeto. No signo a forma pode, ou não, estar incorporada onticamente. Mas ela deve estar incorporada representativamente, ou seja, com respeito à forma comunicada, o signo produz sobre o interpretante um efeito semelhante ao que o próprio objeto produziria sob circunstâncias favoráveis (EP2: 544n.22, 1906).

O que é uma 'forma'? Não pode ser uma forma Platônica, que é essencialmente não comunicável ao menos para algum ser dotado de intuição intelectual, uma faculdade cuja descrição não é encontrada no conceito de *quasi-mente*. Tem que ser algum tipo de entidade que acomode o imediatismo não-intuicional. Tem que ser algo que possa passar do objeto para o signo e do signo para o interpretante enquanto permanece no objeto e no signo. Não pode ser, desse modo, uma 'coisa', isto é, alguma substância primária *à la* Aristóteles. É algo que está incorporado 'onticamente' no objeto, 'representacionalmente' no signo. Do ponto de vista do objeto, a forma é o único modo que ele tem de atrair a atenção para si próprio e, assim, tem que ser alguma característica essencial completamente realizada no objeto. O objeto aqui falado é o que Peirce chama de objeto dinâmico que é aquele que, sendo externo ao signo, nunca é dado imediatamente no signo, mas pode ser sugerido pelo signo através do processo de interpretação. O objeto dinâmico, como bem mostrou Hulswit, exerce três diferentes tipos de 'influência' no signo, dependendo da natureza do último. Se o signo é icônico, o objeto que o determina é uma 'condição necessária' dele; se o signo é indexical, o objeto que o determina age nele como uma causa eficiente; e se o signo é um símbolo, o objeto que o determina é uma causa final dele (Hulswit 1998: 161-167). A forma que é assim transmitida do objeto para o signo pode tomar diferentes aparências, quer seja o objeto uma possibilidade, uma ocorrência ou uma necessidade condicional. Vamos considerar um exemplo.

O rastro deixado por um cervo na neve é um signo que contém tanto o elemento icônico como o indexical. Como um ícone, ele reproduz a forma inversa do casco das patas do cervo com grande fidelidade, tal que a forma real do casco é uma condição necessária da forma deixada na neve. Como um índice, o rastro é o efeito físico da passagem do cervo pela neve. O índice mantém todos os tipos de elementos, como frescor, tamanho, profundidade, precisão, partes distinguíveis que indicarão a um caçador experiente a informação preciosa sobre a idade do animal, peso, sexo, espécie, comportamento habitual, destino e paradeiro prováveis. Para o observador inexperiente, o rastro simplesmente indicará a recente presença de algum animal com cascos naquela área particular. O signo será então mais icônico, mas especialmente mais indexical (visto que é, como um índice, que o trajeto do casco é semioticamente mais potente), para o caçador experiente do que para o observador inexperiente que pode somente reconhecer a

forma vaga de um casco sem ser capaz de identificar sua origem com qualquer precisão. Ícones, segundo Peirce (EP2: 8), trazem com eles uma capacidade para a experiência, mas esta *capacidade* só pode ser explorada dentro dos limites da experiência real que o intérprete já teve do mundo no qual o signo aparece. A indexicalidade do rastro aumenta com a experiência do intérprete.

O mesmo ocorre com as fotografias: o retrato ou paisagem que elas representam só podem ser reconhecidas por pessoas que tiveram a experiência requerida. Isto não quer dizer que a indexicalidade sempre requer uma experiência mais sofisticada (observação colateral mais rica, para usar uma noção de Peirce) do que a iconicidade para exercer seu poder mais efetivamente no interpretante. Para serem reconhecidos, ícones freqüentemente exigem um considerável, sutil e flexível, poder de discriminação e, também, na medida em que alguém jamais tenha visto um casco de pata em sua vida, jamais poderá conectar o rastro na neve com o casco. Assim, a iconicidade também aumenta com experiência (como faz a simbolicidade). A experiência tenderá a aguçar o reconhecimento do signo, a aumentar a apresentação do signo de modo que, uma vez encontrado, comece a falar conosco, não só tagarelamente, mas com mais precisão. Entretanto, como afirma Peirce, o processo representacional começa com a iconicidade: não há qualquer signo que não incorpore, minimamente, ícones na sua composição. Um rastro de casco, como tal, nunca indicará qualquer coisa se não for identificado primeiro como um rastro de casco (desconsiderando o vocabulário, é claro, que é um caso simbólico). Para um ignorante que nunca viu um casco, o rastro pode simplesmente ser uma série de buracos na neve que, como índice, indica que algo deve ter causado o seu aparecimento. Mas antes de poder ser inferido, o reconhecimento dos buracos como buracos deve ter acontecido primeiro.

Se a aprendizagem é uma questão de aumento da habilidade do sujeito para compreensão de signos, isto começa com a habilidade para compreender ícones. Um índice sem um ícone é cego, um símbolo sem um índice é vazio. Puros índices e puros símbolos não ocorrem, exceto de acordo com a abstrata classificação da teoria semiótica, onde são convenientes seus isolamentos. Como vimos acima, 'Em relação à Forma comunicada, o signo produz sobre o interpretante um efeito semelhante àquele que o próprio objeto, ele próprio, produziria sob circunstâncias favoráveis'. Circunstânci-

as favoráveis levariam à aparição direta do próprio cervo ao observador, por exemplo. Feito o rastro do casco, *ecce cervus*. Novamente, somente um caçador experiente será capaz de dizer a idade do animal, sua espécie, etc. O andarilho inexperiente só poderá ser capaz de dizer que 'há um cervo' ou, se atormentado com a sua ignorância, 'há um animal com cascos'. Assim, o signo produz sobre o interpretante um efeito 'similar' ao da aparição potencial do objeto para o qual ele está. A própria aparição do objeto é, claro, muito mais reveladora do que o rastro do casco, de forma que o efeito é somente 'similar', não 'idêntico' (uma das razões pela qual levamos crianças ao jardim zoológico). A 'forma' não é o próprio animal, exposto. É algo que pode-se 'mover' do cervo para o casco da pata para a mente do intérprete. Sua matéria é assim como uma idéia, uma 'idéia-potencialidade' dotada com um duplo poder de crescimento e incorporação (EP2: 388).

Peirce nos diz que o ser da forma consiste na verdade de uma proposição condicional. O rastro da pata nos conta 'sob circunstâncias favoráveis, que você seria capaz de ver um penta-casco'. Ele pode nos contar mais ou menos, dependendo de nossa familiaridade com o signo (nossa 'experiência colateral' como Peirce diz). A idéia pode então crescer, na proporção da qualidade e da riqueza de sua interpretação. As 'circunstâncias favoráveis' têm tanto a ver com a arte de caminhar silenciosamente, contra o vento, ser acompanhado por um caçador experiente, ou ter estudado livros relevantes sobre cervos. O que é significativo é que um signo carrega primeiramente uma experiência potencial, a fonte que vem do objeto dinâmico — e deixe-nos lembrar que *dunamis* significa poder, no sentido de uma fonte de atualização de eventos. O objeto determina o signo fazendo dele um portador deste poder, como manifestado na proposição condicional. O ser da forma é uma questão da verdade, o que significa que o objeto por trás deve *realmente* carregá-la, em primeiro lugar. Vamos imaginar que os rastros dos cascos são falsos: algum brincalhão, caminhando, os plantou lá com a intenção de imitar um cervo. Mesmo com toda sua experiência, nosso bom caçador é enganado e levado a acreditar que um cervo estava caminhando por lá, há alguns minutos, e está provavelmente se escondendo nos arbustos. Seguindo os rastros até o fim, ele fica surpreso em descobrir uma nova espécie de mamífero de patas rindo dele! Não menos indignado, nosso bom caçador abraça o prático brincalhão e lhe agradece profundamente: 'Quan-

to eu aprendi graças a seu bom truque! Toda minha experiência não me preparou para isto, mas agora eu estou feliz de ver minha experiência grandemente ampliada com esta notável adição às minhas possibilidades de interpretação. Isso funcionou tão bem que eu estou terrivelmente tentado a lhe transformar em um animal empalhado — um espécime raro a ser acrescentado à minha coleção colateral!'

A forma incorporada no brincalhão não era a forma que o caçador foi levado a esperar pela sua interpretação habitual do signo, mas isso ocorreu somente por causa da sua ignorância dos modos enganosos do mundo. A proposição condicional não mostrou ser falsa, mas simplesmente precisava de alguma revisão: 'sob circunstâncias favoráveis, você poderia ver um cervo ou alguma outra coisa capaz de deixar o mesmo tipo de rastro'. Assim a idéia-potencialidade cresce. Aprendizagem, portanto, é uma questão de aumento do campo de interpretação através do teste da experiência. Assim que um teste força sobre nós uma nova interpretação, esta interpretação, uma vez completada, se torna parte de nossa 'experiência colateral' e pode servir para aumentar o poder de um signo. Uma razão pela qual nós nunca nos cansamos de reler bons trabalhos é que, a cada leitura, continuamos a experimentar a vida em toda a sua variedade e a cada experiência aumentamos nossa sensibilidade aos signos. Sensibilidade aumentada significa interpretabilidade aumentada, e vice-versa. Potencialmente, não há nenhum limite para esse processo.

Assim, nós podemos começar a ver agora o que chamamos aprendizagem; vaga mas seguramente, ela deve estar conectada, com qualquer aparência que possa ter, a uma 'semiótica' crescente (de maneira geral), a uma crescente abertura para todo tipo de signos, não só do ponto de vista do seu reconhecimento e interpretação, mas também do ponto de vista da própria criação e refinamento. Porque dentro da relação sígnica os interpretantes têm o poder para re-formar os signos que os determinaram, tanto para preservar quanto para intensificar esta determinação por causa do objeto.

Ora, uma vez que um objeto dinâmico infectou um signo com a sua forma, como aquela forma se move do signo para o interpretante? 'O signo não apenas determina o interpretante a representar (ou para tomar a forma de) o *objeto*, mas também determina o interpretante para representar o signo' (EP2: 477-78).

Como mostra Hulswit, a determinação que o signo exerce sobre o interpretante é parecida com aquela da causa eficiente. Este é o caso se o signo é icônico, indexical ou simbólico. O interpretante é afetado pelo signo e isso carrega uma idéia de irresistível força na conexão. Não vamos esquecer que a relação do signo, como um todo, é irredutivelmente triádica, e que um signo não é um signo se não há, *ipso facto*, um resultado inevitável, um apelo direto para um interpretante. O signo, porém, 'determina *que* (sob circunstâncias favoráveis) um interpretante será criado, mas não determina *qual* interpretante será. O *que* o interpretante será, deverá ser determinado pela causa final do processo da semiose' (Hulswit 1998: 165). Um interpretante é um signo em uma relação triádica com o signo que o gerou. Nesta relação, o interpretante deve (1) admitir o recebimento da forma originada no objeto dinâmico, (2) reconhecer que esta forma, como recebida, assumiu uma certa forma representacional forçada (sobre ele) pela mediação do signo e (3) acrescentar àquela forma um signo de reconhecimento, quer dizer, saber que a forma, como recebida, não é estranha ao interpretante, mas, pelo contrário, já familiar a ele de um modo ou de outro. Isto é crucial: um interpretante que não tem familiaridade com a forma que atravessou a relação triádica, estaria fora de lugar naquela relação e não cumpriria sua própria função de significar. É parte do interpretante *qua* interpretante ter a competência requerida para continuar o processo semiótico. A competência só chega a sua capacidade para conectar a forma, como recebida, a outras instanciações comparáveis a esta forma, instanciações já identificadas e cujas identidades estabelecidas irão, de um lado, permitir que o reconhecimento aconteça e, de outro lado, serão envolvidos pela nova experiência provocada na interpretação. Como funciona o processo de reconhecimento? Ele varia de acordo com a natureza da forma que está sendo transmitida e, assim, também de acordo com a própria forma de transmissão.

Shakespeare, no começo do último ato do 'Sonho de uma noite de verão', fez Theseus pronunciar versos da mais alta importância semiótica: 'E, como corpos de imaginação avantes | As formas das coisas desconhecidas, a caneta do poeta | os transforma em formas e dá para o aéreo nada | uma habitação local e um nome'. Traduzido para o jargão semiótico, com uma pequena adição ao fim: e como a fábrica semiótica de manufatura de signos especialmente apropriados para incorporar as formas transmitidas pelos objetos dinâmicos, o poder de interpretação ocorre devido a noti-

ficação e geração de interpretantes capazes: (1) de associar estes signos com um tipo certo de experiência colateral, (2) de descobrir a identidade das formas que estão sendo transmitidas e (3) de transmitir, por sua vez, estas formas reconhecidas a interpretantes futuros.

Signos transmitem formas e formas são a única chance do objeto se manifestar, atrair atenção e entrar no domínio do conhecimento. A aprendizagem é, em grande parte, uma questão de apreensão de tais formas, ao ser capaz de lhes dar uma 'habitação local', isto é, de descobrir como elas se relacionam com uma dada experiência e, então, lhes 'nomear', incorporando-as em novos signos que façam mais justiça às formas iniciais. A interpretação consiste precisamente neste tipo de atividade contínua: encontrar e/ou planejar signos cujos corpos dão às formas transmitidas uma manifestação sempre crescente, sempre por causa do objeto dinâmico original — o poder que mantém alimentado todo o processo de determinação semiótica. Como sugerido, há três tipos de determinação: condição necessária, causação eficiente e causação final. A estes correspondem três tipos de formas transmissíveis — monádica, diádica e triádica —, e três tipos de signos portadores — icônico, indexical e simbólico.[6] Portanto, pode-se imaginar que há, pelo menos, três tipos gerais de proposições condicionais que o processo de semiose continua proferindo. Para Peirce, a aprendizagem também tem a ver com a apreensão da verdade de tais proposições.

Aprendizagem e representação

'Aprendizagem é representação e, portanto, um outro nome para Terceiridade'. Nossa discussão da quarta asserção já considerou a maioria do que está contido nesta asserção. Dizer que aprendizagem é Terceiridade é fazer uma clara afirmação metafísica — sobre a estrutura da realidade. Isto é uma poderosa generalização e uma conclusão lógica do que foi dito. O que é a aprendizagem? Nenhuma resposta psicológica fará justiça à esta pergunta. É preciso cavar mais fundo. A semiótica pode nos ajudar a descobrir muito sobre suas conseqüências. Mas, mesmo um discurso semiótico sobre a aprendizagem, talvez geral e de difícil alcance, não fará justiça a ela. A aprendizagem, nos fala Peirce, é um outro nome para a terceira das três categorias do pensamento e da natureza. Se a chamamos de representação, mediação, continuidade, cresci-

mento, evolução, nós sempre estaremos lidando com aspectos ligeiramente diferentes, mas interdependentes, da mesma dimensão da realidade. Se a aprendizagem é uma parte intrínseca de nossa vida humana, é porque ela é, em primeiro lugar, uma dimensão intrínseca do próprio universo.

'O homem é um signo': nós somos fundamentalmente seres semióticos. A semiose define a nossa essência e, assim, nós aprendemos, e nossa aprendizagem é, por sua vez, uma emanação da própria aprendizagem do universo. Suas verdades eternas são eternas porque nunca terminam de se moldar, o que faz por determinar — ou filtrar — os signos que nós, entre outros, somos então inclinados a aprender a ler. E, conforme lemos, nos mantemos folheando páginas de um livro do qual compartilhamos a autoria, mas não a última.

Notas

[1] Que isto é um papel de 'mediação', já foi contestado com base na definição de Peirce do signo como aquilo que é determinado por um objeto de modo a determinar um interpretante, a se referir ao mesmo objeto, tal que é o signo que media, e não o interpretante. Parte de minha resposta a esta objeção pode ser encontrada em meu artigo 'Peirce's semiotic monism' (1992), em que mostro que cada um dos três termos da relação sígnica (signo, objeto, interpretante) media os outros dois, embora cada um deles o faça de maneira distinta. Uma relação sígnica (*in abstracto*), é uma genuína relação triádica, para Peirce, e portanto, por definição, isto implica que cada termo da relação é um terceiro e, assim, é da natureza de um mediador. Apesar disso, em 'On a new list of categories' o interpretante deve ser o mediador porque sem ele o predicado nunca poderia ser um signo do sujeito: a cópula que une os dois não pode ser formulada antes que o interpretante tenha feito seu trabalho de comparação e reconhecimento entre o sujeito desconhecido e o conhecido correlato.

[2] É preciso lembrarmos aqui os princípios de sua classificação das ciências, cuja construção não é arbitrária (ver EP2: 258-262).

[3] Uma idéia interessante, que pode ter implicações para os psicólogos, é que a conexão, ou 'o fundir' das idéias em idéias mais gerais, manifesta-se na forma de uma 'sensação viva'. A consciên-

cia, portanto, deve aparecer em um nível elementar, mas geral, uma vez que a aprendizagem é um atributo essencial do que é 'quasi-mental'. Uma sensação viva que emerge é portanto uma quase-consciência que sustenta idéias, formas, que se fundem por possuírem alguma coisa essencial em comum. Seja o que for que as mantém conexas, esse algo tem 'consistência' (no sentido etimológico), e é um princípio da semiótica de Peirce que 'consistência' seja uma marca da representação em funcionamento.

[4] O uso que Peirce faz aqui da palavra 'virtual' está em par com o modificador 'quase', quando usado com 'mente' ou 'ego'.

[5] Com relação à questão sobre se a conclusão é parte do argumento, Peirce afirma o seguinte: 'Com relação a uma outra proposição, chamada de Conclusão, freqüentemente colocada para (talvez necessariamente) completar o Argumento, esta representa plenamente o interpretante e, do mesmo modo, tem uma força peculiar ou relação com o interpretante. Há uma divergência de opiniões entre os lógicos se ela faz parte do Argumento ou não; e, mesmo que tais opiniões não tenham se originado de uma análise exata da essência do Argumento, elas devem ter algum peso. O presente autor, sem estar absolutamente confidente, está muito inclinado a pensar que a Conclusão, apesar de representar o interpretante, é essencial para a plena expressão do Argumento' (CP 2.253).

[6] Propositadamente estou simplificando, dado o espaço limitado. A classificação de categorias dos signos de Peirce permite uma análise muito mais sutil da qual eu posso aqui somente sugerir uma direção geral. O leitor irá perdoar esta necessária colherada para Cerberus.

Referências

DE TIENNE, André. 1992. Peirce's semiotic monism. Em: Signs of Humanity - L'Homme et ses signes (Proceedings of the Fourth Congress of the International Association for Semiotic Studies, general editor Gérard Deledalle), Michel Balat e Janice Deledalle-Rhodes (eds.), Volume 3, Semiotics in the World - La Sémiotique dans le monde. Berlin: Mouton de Gruyter, Approaches to Semiotics, pp. 1291-98.

__. 1996. L'analytique de la représentation chez Peirce. La genèse de la théorie des catégories. Bruxelles: Publications des Facultés Universitaires Saint-Louis.

HOUSER, Nathan. 1987. Toward a Peircean semiotic theory of learning. The American Journal of Semiotics 5 (2): 251-274.

HULSWIT, Menno. 1998. A Semeiotic Account of Causation. The 'Cement of the Universe' from a Peircean Perspective. Tese de doutorado, Katholieke Universiteit Nijmegen, Nijmegen. [Alguns capítulos foram publicados separadamente em Transactions of the Charles S. Peirce Society].

PEIRCE, Charles S. (CP). Collected Papers of Charles Sanders Peirce vols. 1-6, C. Hartshorne e P. Weiss (eds.), 1931-35; vols. 7-8,A. W. Burks (ed.), 1958). Cambridge, Mass.: Harvard University Press. [citações de acordo com volume e parágrafo]

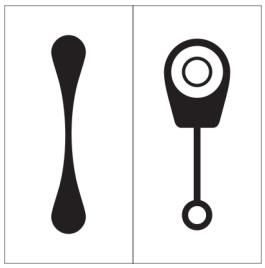
__. (EP1, EP2). The Essential Peirce. Selected Philosophical Writings. vol. 1, Nathan Houser e Christian Kloesel (eds.), 1867-1893; vol. 2, the Peirce Edition Project (ed.), 1893-1913. Bloomington e Indianapolis: Indiana University Press, 1992 e 1998.

__. (W). Writings of Charles S. Peirce. A Chronological Edition. vols. 1-5, Peirce Edition Project (ed.), 1982-1994. Bloomington e Indianapolis: Indiana University Press.

COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA



CAPÍTULO 3

ESTRUTURALISMO HIERÁRQUICO, SEMIOSE E EMERGÊNCIA[1]

Charbel Niño El-Hani e João Queiroz

INTRODUÇÃO

Nosso propósito neste capítulo é discutir em que sentido a semiose pode ser caracterizada como um processo 'emergente'. Este problema foi formulado em projetos sobre simulação computacional, em virtude do largo emprego da noção de emergência. Seu significado, entretanto, raramente é discutido nas ciências da complexidade. Um tratamento rigoroso dessa questão é um dos principais objetivos deste artigo. Antes de começar, devemos esclarecer que não pretendemos responder *quando* ou *como* a semiose emergiu, em termos evolutivos. Estamos mais interessados em discutir as condições que precisam ser satisfeitas para que a semiose possa ser caracterizada como um processo emergente. A solução deste problema é um requisito para a formulação precisa do problema da emergência da semiose em termos evolutivos.

Na próxima seção, veremos como o conceito de emergência tem sido usado nas ciências da complexidade, com atenção ao trata-

mento de Cariani (1989, 1991, 1997). Em seguida, sumarizaremos a análise sistemática das teorias sobre emergência elaborada por Stephan (1998, 1999). No curso desta análise, formularemos questões que devem ser respondidas de modo a precisar a noção de emergência no domínio dos fenômenos semióticos. Em seguida, utilizaremos o estruturalismo hierárquico de Salthe (1987) como base para elaboração de um modelo capaz de explicar 'emergência de semiose' em sistemas que produzem, processam e interpretam signos.

O CONCEITO DE EMERGÊNCIA NAS CIÊNCIAS DA COMPLEXIDADE

O termo 'emergência' (e derivados) tem sido largamente usado em diversos campos de pesquisa, como Vida Artificial e Robótica Cognitiva. Contudo, pouca discussão é encontrada sobre o significado preciso de 'emergência', 'emergente' etc. nestes campos, embora segmentos destes campos cheguem a ser descritos como 'computação emergente' (Cariani 1989, 1991; Emmeche 1994, 1997; Ronald et al. 1999; Bedau 2002; El-Hani 2002). Tendo em vista os debates e as confusões sobre o tema, ao longo do século XX (Blitz 1992, Stephan 1999), é fundamental ter clareza sobre o conceito. A noção de emergência empregada em 'computação emergente' é tão vaga que chegamos a encontrar propostas como a de Ronald (et al. 1999), sugerindo que uma reação de 'surpresa', da parte de um programador, poderia constituir um teste para emergência, em uma simulação computacional. Esta proposta deixa a questão em aberto. Não poderia ser o caso de que o teste sugerido indicasse que não há qualquer emergência nas simulações propriamente ditas? Ele não estaria mostrando que o que ocorre, quando observadores externos supostamente vêem padrões de nível superior emergindo, não é mais do que uma impressão subjetiva? Para responder a estas questões, é preciso enfrentar um problema: a emergência está nas próprias simulações, no observador, ou em ambos? Este problema exige um entendimento mais sofisticado da emergência, em conexão com conceitos relativos a hierarquias, redução, determinação, etc.

Entretanto, há pesquisadores no campo das ciências da complexidade que têm dedicado atenção à necessidade de um tratamento

cuidadoso do conceito, como Mark Bedau e Peter Cariani. Cariani (1989: 146) entende o conceito de emergência como uma tentativa para lidar com a origem da ordem complexa no mundo e concebe duas maneiras para compreender tal origem: explicações da origem da ordem a partir da ordem e explicações da origem da ordem a partir da ausência de forma (*formlessness*) ou estrutura (*structurelessness*).[2] A 'emergência computacional' está comprometida, segundo Cariani (1989: 147), com uma explicação da origem da ordem a partir da ordem. Isso decorre do fato que simulações computacionais são estritamente formais, podendo gerar ordem apenas no sentido da constituição de novas estruturas formais, séries ou cadeias de símbolos, e seus comportamentos, a partir de outras séries ou cadeias de símbolos, também ordenadas. Desse modo, a ordem macroscópica observada nas simulações surge a partir de uma ordem microscópica preexistente nos algoritmos. Todo comportamento que surge nas simulações computacionais deve, assim, ser redutível às conseqüências lógicas de microcomportamentos não-observados, governados pelas regras expressas nos algoritmos subjacentes. Não espanta que Cariani (1989: 148) baseie sua crítica à emergência computacional na discussão de uma proposição que considera um 'artigo de fé' da 'visão de mundo computacionalista', a de que todos os processos físicos são de natureza determinística podendo ser reduzidos a operações primitivas, similares a computações, em algum nível fundamental, de descrição. Toda revolução intelectual suficientemente poderosa produz, segundo Cariani (1997), seus excessos. A visão de mundo computacionalista foi um dos excessos da revolução computacional do século XX:

> Técnicas poderosas rapidamente se tornam visões de mundo totalitárias, que redefinem o mundo à sua imagem, e o computador digital não foi exceção. O dramático surgimento de teorias formais da computação na década de 1930, rapidamente seguido pelo desenvolvimento de computadores eletrônicos digitais nas décadas subseqüentes, efetivamente reviveu idéias platônicas, eventualmente dando luz a ideologias computacionalistas universalistas, com aspirações amplamente anexionistas (Cariani 1997).

O problema que o determinismo coloca para a computação emergente é que não parece fácil conciliar a idéia de um determinismo microscópico com a aceitação de que há emergência no mundo macroscópico (Klee 1984). A ontologia computacionalista parece implicar, assim, a inexistência de processos e comportamentos emergentes, seja no mundo real, seja em simulações. Porém, a

conclusão alcançada por aqueles que defendem a existência de comportamentos emergentes em simulações computacionais é oposta: assumindo que (i) sistemas físicos exibem comportamentos emergentes e (ii) simulações computacionais são isomórficas a sistemas físicos, eles concluem que, em algum nível, as simulações computacionais também deverão exibir emergência. A dificuldade é que a proposição (i) é objeto de discussão, como atestam os debates sobre o conceito de emergência travados desde o começo do século XX. Uma das dificuldades para fundamentar a existência de propriedades emergentes nos sistemas físicos é a possibilidade de que o universo seja determinístico, ou seja, a idéia assumida por uma ontologia computacionalista.

Para Cariani, aparatos formais-computacionais[3] não podem criar novos primitivos:

> *Qualquer processo que possa ser completamente simulado por meio computacional não gerará novos primitivos.* Simulações computacionais de qualquer tipo, sejam simulações de redes neurais, redes conexionistas, [...] autômatos celulares ou modelos evolutivos, não criarão propriedades que não estavam codificadas na simulação, desde seu início (Cariani 1989: 148; Ênfase no original).

Isso implica que, por mais que os pesquisadores na área da computação emergente procurem retirar o elemento humano do circuito, reduzindo a programação a um mínimo, ainda assim os comportamentos emergentes exibidos não seriam mais que uma decorrência da própria programação. Não seriam verdadeiramente emergentes, na medida em que estariam pré-formados. Assim, para Cariani, a razão pela qual não poderia ocorrer emergência, em sistemas formais, no sentido proposto para sistemas naturais, resulta do fato de que um sistema formal é uma criação da mente *em todos os seus aspectos*. Não há maneira de introduzir propriedades adicionais àquelas encontradas desde o início ou, se não for o caso, dedutíveis de nossas definições por meio de regras de inferência. Não aparecerão ao longo da simulação novas propriedades, ou regras, não especificadas, ou logicamente implicadas, desde o início. Novos primitivos não podem ser gerados em uma simulação computacional porque um aparato formal-computacional não é capaz de criar novas relações sintáticas ou semânticas. Assim, Cariani (1989: 155-156, 171) caracteriza aparatos formais, computacionais ou robóticos, como 'não-emergentes'.

Mas o que dizer dos comportamentos emergentes observados em simulações computacionais? Como caracterizar os comportamentos coletivos que simulações exibem, e que programadores não podem prever? Para responder a estas questões, é preciso tratar da estratégia geral empregada na construção de simulações nas ciências da complexidade (ver Cariani 1989: 149; Emmeche 1997). Esta estratégia consiste em desenvolver uma simulação com tal número de primitivos e regras de interação que se torna difícil para um observador externo (incluindo o programador) prever o comportamento da simulação. Quando a simulação é executada, o observador encontra padrões inesperados. Afirma-se que tais padrões *literalmente* emergiram no curso da simulação. O comportamento emergente é atribuído ao próprio aparato formal-computacional, à própria simulação.

É neste ponto que uma questão se torna inevitável: não seria mais natural pensar que se trata de um processo de emergência aos olhos do observador, tornando difícil atribuir a emergência à própria simulação? Esta idéia é defendida por Cariani:

> Poder-se-ia argumentar que, no curso da simulação de um autômato celular, ou de uma simulação evolutiva, 'padrões de nível superior' emergem, mas estas são distinções geradas pelo observador humano e não pela própria simulação (Cariani 1989: 172; tb. 176).

Contudo, qualquer processo de emergência, incluindo aqueles observados em sistemas naturais, só pode ser entendido como tal se for assim reconhecido por um observador (ver Baas 1996, Emmeche 1997, Baas & Emmeche 1997).[4] Da perspectiva defendida por estes autores, a emergência é entendida como uma *estratégia explicativa*, destacando a função do observador na qualificação de uma propriedade emergente. Baas analisa a emergência em termos de uma série de processos abstratos de construção. Considere-se um conjunto S_1 de estruturas de primeira ordem. Por algum mecanismo observacional $Obs_1(S_1)$, são obtidas, ou medidas, suas propriedades. Estas estruturas podem estar sujeitas a uma família de interações, *Int*, a partir das quais um novo tipo de estrutura aparece, $S_2 = R\ (S_1, Obs_1(S_1), Int)$, onde *R* corresponde ao resultado do processo de construção. S_2 é uma estrutura de segunda ordem, cujas propriedades podem ser obtidas por meio de outro mecanismo observacional, Obs_2, que também é capaz de observar as estruturas de primeira ordem. Baas (1996) define *P* como uma proprie-

dade emergente de S_2, se, e somente se, *P* pertence ao conjunto $Obs_2(S_2)$, mas não ao conjunto $Obs_2(S_1)$. Esta definição mostra como propriedades e processos emergentes apenas podem ser entendidos como tais de uma perspectiva observacional bem definida. Ela, no entanto, pode ser considerada incompleta, na medida em que pontos centrais não são contemplados, a exemplo da relação que se supõe existir entre propriedades e processos emergentes e a microestrutura dos sistemas que os exibem (El-Hani 2002).

A despeito da dependência do observador ser uma característica compartilhada por emergência computacional e sistemas naturais, pode-se conceber uma diferença no tipo de dependência do observador. Mas qual a diferença? Não se trata simplesmente do fato que padrões de nível superior, que emergem nas simulações, devem ser reconhecidos por um observador humano, visto que isso se aplica a qualquer tipo de padrão, incluindo aqueles observados em qualquer sistema natural. A questão reside na relação determinística entre as regras codificadas no algoritmo, no qual está baseada a simulação, e os comportamentos exibidos. Não surgem, ao longo da simulação, novas regras. Em um sentido preciso, todas as regras estão pré-especificadas no algoritmo. Desse modo, não pode surgir na simulação qualquer comportamento que não seja uma conseqüência lógica das regras e do estado inicial da simulação (Cariani 1989: 149). A evolução temporal de um autômato celular, por exemplo, deve sempre começar a partir de um conjunto inicial de operadores e operantes primitivos, e quaisquer outros que surjam ao longo da simulação devem ser redutíveis a *combinações* deste conjunto inicial. Não podem, portanto, ser considerados novos primitivos, conforme definido por Cariani.[5]

Embora tenha reconhecido que a combinação de primitivos pode resultar no surgimento de padrões inesperados, de 'comportamentos não-antecipados com categorias completamente antecipadas' (Cariani 1989: 171), Cariani propôs, posteriormente, uma forma de emergência em tais simulações. Em 1997, ele definiu duas concepções complementares de emergência: a emergência 'combinatorial', na qual a novidade surge como resultado de novas combinações de primitivos preexistentes, consistindo em uma produção de ordem a partir de ordem; e a emergência 'criativa', na qual a novidade surge por meio da criação *de novo* de tipos inteiramente novos de primitivos, correspondendo à produção de ordem a partir de ruído, caos ou ausência de forma (*formlessness*). Estas

duas concepções de emergência propiciam dois modos de descrever e compreender a mudança: no primeiro caso, a mudança consiste no desdobramento de conseqüências de um conjunto fixo de regras, enquanto, no segundo, a mudança consiste no surgimento de novos processos e novos tipos de interações ao longo do tempo. A novidade, no primeiro caso, nada mais é que o 'desdobramento preformacionista da possibilidade latente' (Cariani 1997). A emergência criativa, por sua vez, está relacionada à criação de novos padrões, à formação de novos domínios de possibilidades. Apesar de propor uma forma de emergência em simulações computacionais, Cariani mantém a idéia de que há uma diferença fundamental entre a emergência em sistemas biológicos e simulações computacionais. Nestas últimas, somente a emergência de natureza combinatorial seria possível, dado que a preexistência de um conjunto de primitivos torna impossível a ocorrência de emergência criativa, i.e., o surgimento de primitivos novos, como ocorre, para Cariani, na emergência biológica.[6]

Os argumentos de Cariani (1989), sobre o que seria necessário para obter emergência criativa em simulações computacionais (e.g. autômatos celulares), indicam a necessidade de irmos além da virtualidade, construindo entidades que apresentem uma estrutura física concreta e estejam situadas em um ambiente externo:

> Que tipos de comportamentos seriam necessários para gerar emergência fundamental em um computador? São necessárias *dinâmicas* de interação *dependentes de concentração* (*rate-dependent*), fisicamente coerentes, *que não sejam especificadas explicitamente* pelo programador desde o início, *que poderiam entrar em jogo espontaneamente* durante o curso de uma simulação. Estas teriam de ser regras ou dinâmicas que poderiam mudar sem invocarem uma outra regra pré-especificada. Um exemplo disso teria lugar se o comportamento global de um autômato celular começasse a modificar ou a restringir as regras locais de interação, sem que as interações global-local tenham sido especificadas pelo programador da simulação desde o início (Cariani 1989: 176; Ênfase no original).

Mais adiante, ele é ainda mais explícito sobre a necessidade de corpos e ambientes reais:

> [...] se novas interações, e comportamento em aberto (*open-ended*), são desejados, processos de redes físicas, análogos, não podem ser somente simulados, eles devem ser implementados fisicamente, ou não entrarão em jogo quaisquer novas interações (Cariani 1989: 177).

Assim, a emergência criativa exige novas abordagens. Elas são representadas, hoje, por duas tendências, ambas com forte ênfase no papel do *corpo* do agente e do *ambiente* em que estão imersos: cognição corporificada (*embodied cognition*) e situada (*situated cognition*) (ver Clark 1997, 2001).

Em vez de aparatos que operam com base em categorias pré-especificadas, a emergência criativa exige, argumenta Cariani (1997), a produção de aparatos que tenham seus próprios meios para alterar adaptativamente suas estruturas internas. Estes aparatos deverão ser capazes de construir seus próprios primitivos, seus próprios critérios de relevância para a avaliação das experiências e de suas conseqüências adaptativas no ambiente onde estão situados. Em suma, para produzir eventos de emergência criativa, é preciso ir além do aspecto formal das simulações computacionais, construindo aparatos que apresentam um aspecto material e, portanto, deixando de lado a visão funcionalista pura que caracterizou a maior parte da computação emergente. Como argumenta Cariani (1989), 'nós podemos ter aparatos emergentes se abandonarmos a natureza determinística, simbólica, dos aparatos, e podemos ter simulações computacionais determinísticas, bem comportadas, desde que abandonemos a esperança de torná-las emergentes, mas não podemos ter ambos ao mesmo tempo' (Cariani 1989: 151). Este ponto não tem sido freqüentemente reconhecido pela IA, uma vez que, em contraste com o reconhecimento do poder generativo da novidade combinatorial, pesquisadores não dão a devida atenção à necessidade de processos que criem novos primitivos (Cariani 1997).

No caso de entidades realizadas materialmente nas quais novos primitivos e novas regras de interação irredutíveis a regras pré-especificadas podem surgir, o sistema em questão não pode mais ser definido como um sistema formal realizando computações. No caso de entidades corpóreas (*embodied*) e situadas, como aquelas construídas em robótica evolutiva, as regras de transição se tornam ambíguas, as funções de *input-output* se tornam indeterminadas. As características definidoras da natureza formal de um sistema são perdidas. Trata-se de uma situação em aberto (*open-ended*) na qual a novidade pode surgir dos padrões de interação da entidade realizada materialmente com seu ambiente.

Cariani (1989, 1997) contrapõe à emergência combinatorial, em sistemas formais-computacionais, um conceito derivado de Robert Rosen de emergência relativa a um modelo, em sistemas naturais.

De acordo com Rosen, a emergência é um desvio do comportamento de um sistema natural em relação a um modelo do mesmo. Este conceito de emergência tem, contudo, uma natureza estritamente epistêmica. Trata-se do mesmo tipo de interpretação que encontramos em Nagel (1961: 369):

> É claro [...] que dizer de uma dada propriedade que ela é 'emergente' significa atribuir-lhe um caráter que a propriedade pode possuir em relação a uma teoria ou um corpo de suposições, mas pode não possuir em relação a alguma outra teoria. Desse modo, a doutrina da emergência [...] deve ser entendida como afirmando certos fatos lógicos acerca de relações formais entre enunciados, e não quaisquer fatos 'metafísicos' acerca de alguns traços supostamente 'inerentes' de propriedades dos objetos.

A compreensão da emergência como uma noção meramente epistêmica empobrece este conceito e, por estas razões, foi recusada pela grande maioria dos pensadores emergentistas.

Como vimos, a dependência do observador não é uma exclusividade de processos computacionais emergentes. Processos emergentes, em sistemas naturais, também devem ser percebidos pelo observador como tais. Resta, assim, o determinismo característico do aparato formal-computacional como dificuldade para a tese de que há comportamentos emergentes em simulações computacionais. Nas próximas seções, apresentaremos uma caracterização do emergentismo que não é avessa a idéia de determinação, incluindo uma tese de determinação sincrônica das propriedades e dos processos emergentes, e, ao menos para autores comprometidos com o emergentismo britânico clássico, uma tese de determinação diacrônica. O que está em questão: que outras características os processos devem exibir, em sistemas naturais ou em simulações computacionais, para que possam ser qualificados como emergentes?

Características centrais do emergentismo e algumas questões sobre a semiose

A semiose pode ser descrita como um processo 'emergente' em sistemas semióticos. Mas qual o significado preciso desta descrição? Esta questão se tornou particularmente importante na última década, na qual foi intensificado o debate sobre a emergência (ver Kim 1998, 1999; Stephan 1999; Cunningham 2001; Pihlström 2002;

El-Hani 2002). O emprego crescente da idéia de emergência, não apenas na computação emergente, mas também nas neurociências, ciências cognitivas, biologia de sistemas, genômica etc., torna crucial evitar sua aplicação de maneira imprecisa, sobretudo por tratar-se de um conceito que teve sua história marcada por grande confusão, no que diz respeito aos seus aspectos metafísicos e epistemológicos.

Nosso propósito aqui é explorar uma aplicação precisa do conceito, tanto em sistemas naturais, quanto em ambientes de simulação computacional. Colocaremos em diálogo duas tradições filosóficas, a semiótica e o emergentismo. Empregaremos a análise sistemática de teorias da emergência desenvolvida por Stephan (1998, 1999). Ao apresentá-la, formularemos questões que devem ser respondidas para que a noção de emergência seja utilizada de modo preciso.

Freqüentemente, entende-se 'emergência' de uma maneira intuitiva, como 'criação de novas propriedades'. Esta definição remete-se a uma das fontes do pensamento emergentista, a obra do psicólogo britânico Conwy Lloyd Morgan. Como Emmeche e colaboradores (1997) mostraram, uma discussão dos conceitos envolvidos ('novidade', 'propriedade' e 'criação') pode levar a alguma compreensão dos principais tópicos do pensamento emergentista. Contudo, esta definição não é suficiente para um entendimento preciso do conceito de emergência, principalmente porque concentra-se em idéias características de um tipo de emergentismo, o 'emergentismo diacrônico' (ver abaixo).

Em um sentido técnico, 'propriedades emergentes' podem ser entendidas como uma certa classe de propriedades de nível superior que se relacionam *de uma certa maneira* à micro-estrutura de *uma classe de sistemas*. A razão pela qual uma definição assim tão ampla, com tantas cláusulas em aberto (indicadas em itálico), parece mais adequada tem a ver com o fato de que o conceito de emergência é usado em diversos campos. É provável que uma definição mais detalhada se aplique a um campo específico. É evidente que uma definição mais operacional é necessária, em casos particulares. Nossa idéia é que devemos torná-la mais precisa, considerando restrições teóricas e empíricas específicas. É parte da tarefa de uma teoria da emergência, aplicada a um campo particular, preencher as cláusulas em aberto da definição acima. Uma teoria da emergência deve, entre outras coisas, prover uma caracterização de quais propriedades de uma classe de sistemas devem ser consi-

deradas 'emergentes', e uma explicação sobre como estas propriedades se relacionam com a micro-estrutura de tais sistemas. Ela deve também estabelecer a classe de sistemas que exibem certo conjunto de propriedades emergentes.

Ao estender a definição acima, para que ela se refira não apenas a propriedades, mas também a processos, surge a primeira questão para a caracterização da semiose como um processo emergente. Ela se refere a classe de sistemas que exibem semiose: (1) o que é um sistema semiótico?

Não existe uma teoria unificada da emergência. Contudo, é possível reconhecer, entre as teorias, uma série de características centrais (Stephan 1999, capítulo 3). Primeiramente, emergentistas devem estar comprometidos com o *naturalismo*, assumindo que apenas fatores naturais exercem um papel causal no universo. Ainda que naturalismo e materialismo não coincidam filosoficamente, é o caso que um emergentista naturalisticamente orientado deve também comprometer-se com a idéia de que todas as entidades consistem de partes materiais. Esta tese é denominada *monismo físico:* no universo, existem apenas entidades constituídas física ou materialmente, e qualquer propriedade ou processo emergente é instanciado por sistemas constituídos exclusivamente por partes materiais. A questão seguinte é: (2) os sistemas que exibem semiose são constituídos apenas fisicamente?

Uma segunda característica do emergentismo é a noção de novidade: novos sistemas, estruturas, processos, entidades, propriedades e disposições são formadas no curso da evolução. Isso nos leva a questão: (3) sistemas semióticos constituem uma nova classe de sistemas, instanciando novas estruturas, processos, propriedades, disposições, etc?

As teorias da emergência requerem, em terceiro lugar, uma distinção entre *propriedades sistêmicas* e *não-sistêmicas*. Uma propriedade sistêmica é encontrada no nível do sistema e não no nível de suas partes; de outro lado, uma propriedade não-sistêmica é também encontrada nas partes do sistema. Baseados na distinção entre processos sistêmicos e não-sistêmicos: (4) a semiose pode ser descrita como um processo sistêmico?

Uma quarta característica das teorias da emergência é a noção de *hierarquia de níveis de existência*. (5) Como devem ser descritos níveis em um sistema semiótico?

Uma quinta característica é a tese da *determinação sincrônica*, um corolário do monismo físico: as propriedades e disposições comportamentais de um sistema dependem de sua micro-estrutura, isto é, das propriedades e arranjos de suas partes. Não pode haver diferença nas propriedades ou nos processos sistêmicos sem que haja alguma diferença nas propriedades das partes do sistema e/ou em seu arranjo. (6) Em que sentido podemos dizer (e explicar) que a semiose é sincronicamente determinada pelas propriedades e arranjos das partes do sistema?

Em sexto lugar, embora alguns emergentistas (e.g., Popper, em: Popper & Eccles [1977]1986) tenham se comprometido com o indeterminismo, uma das características do emergentismo (ao menos na tradição britânica clássica) é a crença na *determinação diacrônica*: o advento de novas estruturas seria um processo determinístico governado por leis naturais (Stephan 1999: 31). Esta é certamente uma característica das teorias clássicas da emergência que é incompatível com a moldura teórica de Peirce, já que ele rejeita a crença em um universo determinista (CP 6.201). Mas isso não impede o tratamento da emergência em conexão com a abordagem peirceana da semiose, uma vez que existem teorias da emergência comprometidas com o indeterminismo. Não é necessário, de modo algum, prender-se ao emergentismo britânico clássico.

Em sétimo lugar, emergentistas estão comprometidos com a noção de *irredutibilidade* de uma propriedade sistêmica classificada como 'emergente'. Uma oitava noção característica do emergentismo é a de *imprevisibilidade* em princípio. Formulamos aqui duas perguntas: (7) em que sentido podemos dizer que a semiose é irredutível? (8) Em que sentido podemos dizer que a instanciação da semiose em sistemas semióticos é imprevisível em princípio?

Finalmente, a nona característica do emergentismo é a idéia de *causação descendente*: novas estruturas e novos tipos de estados de relação (relatedness) entre objetos pré-existentes manifestam eficácia causal descendente, determinando o comportamento de suas partes. A questão seguinte: (9) Alguma forma de causação descendente poderia estar envolvida na semiose? As noções de imprevisibilidade e irredutibilidade serão discutidas em detalhes na próxima seção e as questões relacionadas a elas serão refinadas.

VARIEDADES DE EMERGENTISMO E ALGUMAS QUESTÕES SOBRE A SEMIOSE

É possível construir uma tipologia das variedades de emergentismo tomando como base as características centrais discutidas acima (Stephan 1999: capítulo 4). Para nossos propósitos vamos considerar apenas três variedades de emergentismo - fraco, sincrônico e diacrônico. O emergentismo fraco assume: (1) monismo físico, (2) distinção entre propriedades sistêmicas e não-sistêmicas, e (3) determinação sincrônica. Tais características constituem as condições mínimas para uma filosofia emergentista materialista. O emergentismo fraco é a base comum para todas as teorias materialistas da emergência mais fortes. Contudo, o emergentismo fraco é compatível com o fisicalismo redutivo (Stephan 1998: 642; 1999: 67) fazendo-o insuficiente face às motivações da maioria dos teóricos da emergência, que consideram o emergentismo uma posição antireducionista. Neste capítulo, caracterizamos a semiose como um processo emergente no sentido de uma teoria da emergência mais forte, sendo necessário analisarmos os conceitos de *irredutibilidade e imprevisibilidade*, assumidos em teorias da emergência sincrônicas e/ou diacrônicas.

Combinando à visão emergentista fraca a tese da irredutibilidade de propriedades ou processos sistêmicos, o emergentismo sincrônico constitui uma doutrina incompatível com o fisicalismo redutivo. Stephan (1998: 642-643; 1999: 68) distingue dois tipos de irredutibilidade. A primeira noção de irredutibilidade é baseada na não-analisabilidade das propriedades sistêmicas: (I_1) [Irredutibilidade como não-analisabilidade] Propriedades sistêmicas que não podem ser analisadas em termos do comportamento das partes de um sistema são necessariamente irredutíveis (cf. Stephan 1998: 643). Esta noção, que cumpre um papel importante nos debates sobre os *qualia*, está relacionada a uma primeira condição de redutibilidade, que uma propriedade *P* será redutível se, do comportamento das partes do sistema, seguir que o sistema exibe *P*. Inversamente, uma propriedade sistêmica *P* de um sistema *S* será irredutível se não seguir, nem mesmo em princípio, do comportamento das partes de *S* que *S* exiba *P*.

A segunda noção de irredutibilidade se baseia na não-dedutibilidade do comportamento das partes do sistema: (I_2) [Irredutibilidade do comportamento das partes do sistema] Uma propriedade sistêmica será irredutível se ela depender do comportamento específico que as partes exibem no interior de um sistema de um dado tipo e este comportamento não seguir do comportamento que os componentes apresentam quando isolados, ou quando no interior de sistemas de tipos mais simples (cf. Stephan 1998: 644). Este conceito de irredutibilidade está relacionado à noção de *causação descendente*: parece haver uma influência causal descendente do sistema no qual uma propriedade, ou um processo, emergente P é observado sobre o comportamento de suas partes, o que impede uma dedução deste último do comportamento que aquelas mesmas partes exibem quando isoladas ou como partes de sistemas de tipos mais simples. Uma segunda condição de redutibilidade é violada neste caso, implicando que uma propriedade sistêmica *P* de um sistema *S* será irredutível se ela for realizada por partes do sistema *S* cujo comportamento não segue, nem mesmo em princípio, do comportamento que elas próprias exibem em sistemas mais simples do que *S*.[7]

Uma análise mais detalhada do conceito de irredutibilidade nos leva a uma reformulação da sétima questão: (7) Qual interpretação da irredutibilidade é mais adequada para tratar a irredutibilidade da semiose: não-analisabilidade ou não-dedutibilidade? Além disso, a explicação da irredutibilidade enquanto não-dedutibilidade, acima, torna evidente que a questão 9 (Alguma forma de causação descendente poderia estar envolvida na semiose?) deve surgir em conexão com esta interpretação específica da irredutibilidade da semiose. Evitaremos o problema da causação descendente aqui, uma vez que ele requer uma discussão complexa sobre 'causalidade'.

O emergentismo diacrônico se ocupa da noção de 'evolução emergente'. Teorias diacrônicas da emergência tratam da tese de que o processo evolutivo resulta em *novidades qualitativas*, opondo-se a qualquer tipo de preformacionismo. Contudo, a simples adição do conceito de novidade não é suficiente para a formulação de uma teoria da emergência forte. É preciso ir além, apoiando a tese de que novas estruturas e propriedades são imprevisíveis por uma questão de princípio (em oposição a uma imprevisibilidade prática). Neste contexto entra em cena a distinção entre uma 'novidade simples' e uma 'novidade genuína', no jargão emergentista. Uma propriedade ou estrutura é considerada genuinamente nova porque

seu aparecimento não poderia ter sido previsto com base em um conhecimento completo sobre o estado do universo. Uma propriedade, ou processo sistêmico, poderia ser imprevisível em princípio por duas razões (Stephan 1998: 645): (i) porque a micro-estrutura do sistema, em que a propriedade ou o processo é instanciado (e que o determina sincronicamente), é imprevisível em princípio; (ii) porque a propriedade ou o processo é irredutível, não importando se a micro-estrutura do sistema é imprevisível em princípio.[8] Este segundo caso não oferece ganhos adicionais, relativamente àqueles obtidos no tratamento da irredutibilidade.Por esta razão, ao discutir o problema da imprevisibilidade em princípio, colocaremos em foco a imprevisibilidade da estrutura de sistemas ou processos semióticos, reformulando a oitava questão (seção anterior): (8) a estrutura de sistemas ou processos semióticos pode ser considerada imprevisível em princípio?

Antes de propor respostas para as questões formuladas, vamos apresentar um modelo baseado no *estruturalismo hierárquico* de Salthe (1985). O modelo foi elaborado e desenvolvido em diversos trabalhos (Queiroz & El-Hani 2006a,b; El-Hani et al. 2006).

Níveis de semiose: um modelo geral

Salthe (1985: 21) propõe, para descrição de sistemas complexos, o que chama de 'estruturalismo hierárquico'. Um elemento fundamental do estruturalismo hierárquico é o 'sistema triádico básico', elaborado sob influência Peirceana.[9] De acordo com o sistema triádico, para descrever as interações fundamentais de uma dada entidade, ou processo, é necessário: (i) considerá-lo no nível em que efetivamente o observamos ('nível focal'); (ii) investigá-lo em termos de suas relações com as partes descritas em um nível inferior; e (iii) considerar as entidades e processos em um nível superior, em que estão *imersos* entidades e processos observados no nível focal. Para Salthe, tanto o nível inferior, quanto o superior, exercem influências restritivas (*constraining*) sobre a dinâmica das entidades e/ou processos no nível focal. Estas restrições permitem explicar a emergência de entidades e processos (e.g. semiose) no nível focal.

No nível inferior, as condições de restrição correspondem a 'potencialidades' ou 'condições iniciadoras' (*initiating conditions*) dos

processos emergentes, enquanto as condições de restrição estabelecidas pelo nível superior estão relacionadas ao papel do ambiente (seletivo). Esta classe de restrições corresponde a 'condições de contorno' que coordenam ou regulam a dinâmica no nível focal.[10]

Os processos que emergem no nível focal são instanciados por meio da interação entre processos que têm lugar em níveis imediatamente superiores e inferiores. Os fenômenos observados no nível focal devem estar '... entre as possibilidades engendradas por permutações de condições iniciadoras possíveis estabelecidas no nível imediatamente inferior' (Salthe 1985: 101). Processos no nível focal estão imersos em um ambiente de nível superior, que seleciona, entre os estados potencialmente engendrados pelos componentes no nível inferior, aqueles que serão efetivamente realizados (*actualized*). A figura 1 mostra um esquema das relações determinativas no sistema triádico.

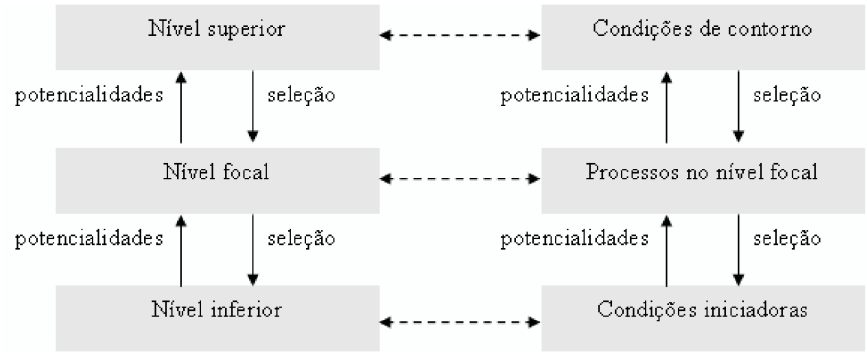


Figura 1: Esquema das relações determinativas no sistema triádico de Salthe. O nível focal não somente é restringido por condições de contorno, como estabelece as potencialidades para a constituição deste. Similarmente, o nível focal não somente é constituído a partir de potencialidades estabelecidas pelo nível inferior, como também estabelece condições de contorno para os processos que têm lugar neste.

Vamos considerar que um determinado processo deve ser localizado no nível em que é efetivamente observado, o 'nível focal'. Processos semióticos no nível focal serão descritos como cadeias de tríades. Podemos discutir a relação entre processos semióticos no nível focal e entidades e/ou processos em um nível inferior ('nível micro-semiótico') e em um nível superior ('nível macro-semiótico'). Neste último, são descritas redes de cadeias de tríades, nas quais os processos semióticos no nível focal estão imersos. No nível micro-semiótico, devem ser consideradas as relações de determinação que têm lugar em cada tríade S-O-I. As relações de determinação

provêem o modo como os elementos (S-O-I) estão arranjados. De acordo com Peirce, o interpretante é determinado pelo objeto através da mediação do signo (MS 318: 81). Isso resulta em duas relações determinativas: a determinação do signo pelo objeto relativamente ao interpretante, e a determinação do interpretante pelo signo relativamente ao objeto (ver De Tienne 1992).

No nível micro-semiótico, deve-se considerar que, dadas as posições relativas de S, O e I, uma tríade ti = (Si, Oi, Ii) somente pode ser definida no contexto de uma cadeia de tríades T = {..., ti-1, ti, ti+1,...} (ver Gomes et al. prelo, 2003a,b). Como Savan (1986: 134) destaca, um interpretante é o terceiro termo de uma relação triádica e o primeiro termo (signo) de uma relação triádica subseqüente. Essa é a razão pela qual a semiose não pode ser definida em uma tríade isolada; ela necessariamente envolve cadeias de tríades (Merrell 1995). Isso é uma indicação de que a semiose pode ser caracterizada como um processo sistêmico, encontrado somente no nível focal, mas não no nível das partes que a compõem, as tríades no nível micro-semiótico (ver questão 4).

A micro-semiose estabelece as condições iniciadoras dos processos no nível focal. Cada cadeia de tríades indica o mesmo objeto dinâmico, através de diversos objetos imediatos, como representados em cada tríade. As possibilidades de indicação de um objeto dinâmico são restringidas pelas relações de determinação em cada tríade. Assim, o modo como *O* determina *S* relativamente a *I*, e *S* determina *I* relativamente a *O*, e, por conseguinte, como *I* é determinado por *O* através de *S* conduz a um número potencial de maneiras pelo qual o objeto dinâmico pode ser indicado na semiose, no nível focal, isto é, a um conjunto de relações triádicas potenciais entre objetos imediatos, signos e interpretantes.

Introduzimos uma distinção entre *potencialidade* e *atualidade*. Um 'signo potencial' é algo que *pode* ser signo de um objeto para um interpretante; um 'objeto potencial' é algo que *pode* ser um objeto de um signo para um interpretante; um 'interpretante potencial é algo que *pode* ser um interpretante de um signo. O nível micro-semiótico pode ser definido como um domínio de signos, objetos e interpretantes potenciais. Devemos considerar um conjunto *W* de possíveis relações determinativas entre estes elementos, que podem gerar um conjunto de possíveis tríades. Estas tríades não podem ser 'fixadas' no nível micro-semiótico, uma vez que este estabelece somente condições iniciadoras para cadeias de tríades

no nível focal. O estabelecimento das cadeias também depende de condições de contorno estabelecidas pelo nível superior.

O nível macro-semiótico estabelece as condições de contorno para a realização dos processos semióticos no nível focal. A influência seletiva das condições de contorno estabelecidas pelas redes de cadeias, que constituem um ambiente ou contexto, é fundamental para a atualização de cadeias de tríades no nível focal. São selecionadas, entre as tríades potencialmente engendradas pelas relações determinativas no nível micro-semiótico, aquelas que serão efetivamente atualizadas. Como vimos, uma tríade $t_i = (S_i, O_i, I_i)$ não pode ser definida atomisticamente, mas somente quando imersa em estruturas (e/ou processos) de nível superior, incluindo tanto cadeias de tríades, $T = \{..., t_{i-1}, t_i, t_{i+1},...\}$, quanto redes de cadeias de tríades, $R = \{T_1, T_2, T_3,..., T_n\}$. Estas estruturas e/ou processos de nível superior provêem o contexto para a atualização de relações determinativas potenciais em cada tríade. Uma cadeia de tríades, $T_i = \{..., t_{i-1}, t_i, t_{i+1},...\}$ é então formada pela atualização, sob a influência regulatória do nível macro-semiótico, de uma série de tríades potenciais engendradas no nível micro-semiótico.

É neste sentido que a emergência de processos semióticos no nível focal, em que cadeias de tríades são atualizadas, pode ser explicada como o resultado da interação entre potencialidades estabelecidas no nível micro-semiótico e a influência seletiva, regulatória, do nível macro-semiótico. As idéias gerais envolvidas neste modelo de semiose em três níveis são mostradas na Figura 2.

Respostas para as questões sobre a semiose

Vamos considerar a questão: (1) o que é um sistema semiótico? A semiose pode ser definida como um processo auto-corretivo envolvendo a cooperação interativa de três componentes, S-O-I. Os sistemas que nos interessam, chamados por Fetzer (1988) de 'sistemas semióticos', podem ser definidos como a incorporação (*embodiment*) de tal processo. Um sistema semiótico é um sistema que produz, transmite, interpreta signos de diferentes tipos. Para Fetzer (1997: 358), o que torna um sistema 'semiótico' é o fato de que ele é 'causalmente afetado pela presença de um signo, porque, para o sistema, ele se refere a alguma outra coisa, icônica, indexical ou simbolicamente. Estas coisas às quais os signos se referem po-

dem incluir objetos e propriedades abstratas, teóricas, não-observáveis, ou não-existentes, que podem ser incapazes de exercer qualquer influência causal sobre um sistema, por si mesmas.' Sistemas semióticos apresentam comportamento auto-corretivo, ou algum tipo de atividade direcionada a um fim (ver Ransdell 1977: 162). Eles são capazes de usar signos como meios para a comunicação de uma forma, ou transferência de um hábito, incorporada no objeto, para o interpretante, ou, no caso de sistemas biológicos, o intérprete, de tal forma a restringir seu comportamento (EP 2:544 n.22; tb. EP 2:391, 2:477).

A segunda questão diz respeito à natureza dos sistemas semióticos: (2) eles são constituídos apenas fisicamente? Processos semióticos podem ser realizados apenas através de implementação física (Ransdell 1977). Portanto, sistemas semióticos devem ser materialmente incorporados (Emmeche 2003, Deacon 1999: 2). Se um signo deve possuir um modo de ser ativo, ele deve ser instanciado materialmente. Peirce considera as qualidades materiais do signo como as características que pertencem ao signo, em si mesmo: 'Como um signo não é idêntico à coisa significada, mas difere dela em alguns aspectos, ele deve claramente possuir algumas características pertencentes a ele próprio, que nada têm a ver com sua função representativa. Eu as chamo de qualidades materiais do signo' (CP 5.287).

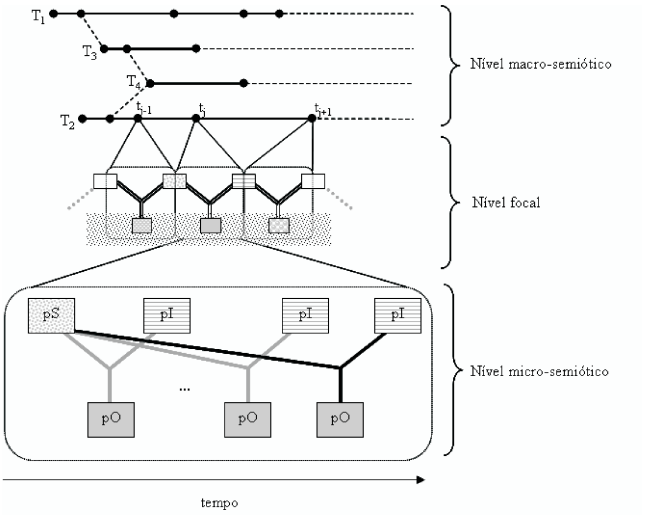


Figura 2: Modelo de semiose em três níveis. pS: signo potencial; pO: objetos potenciais; pI: interpretantes potenciais. A área cinza no nível focal indica que todos os objetos nas tríades de uma cadeia indicam o mesmo objeto Dinâmico.

A terceira questão: (3) sistemas semióticos constituem uma nova classe de sistemas, instanciando novas estruturas, processos, propriedades, disposições etc.? Neste capítulo, não temos qualquer intenção de definir onde se encontra o limiar a partir do qual sistemas semióticos são encontrados na história do universo. Assumimos que houve um período na história do universo no qual sistemas capazes de usar signos não existiam. Portanto, ainda que relações irredutivelmente triádicas possam ter precedido a origem de sistemas semióticos, sugerimos que esta classe de sistemas surgiu em um certo momento da história evolutiva do universo. Postulamos que, antes do aparecimento de sistemas semióticos, existiam sistemas meramente reativos, incapazes de interpretar signos. Certamente, existiam coisas no mundo às quais sistemas materialmente incorporados reagiam, mas tais sistemas não eram capazes de usar tais coisas como signos, como meios para a comunicação de formas, i.e., eles não eram intérpretes. Nada além de uma dinâmica de sistemas e coisas diadicamente acopladas teria existido, sem que qualquer processo interpretativo tivesse lugar. Nesses termos, podemos dizer que sistemas semióticos constituem uma nova classe de sistemas, com um novo tipo de estrutura, capaz de produzir e interpretar signos, e assim, de realizar semiose, como um novo tipo de processo (emergente).

À primeira vista, pode parece incompatível a idéia de que sistemas semióticos constituem uma nova classe de sistemas com a tese metafísica sinequista de Peirce, dada a compreensão típica da doutrina da emergência como sendo comprometida com a tese de que a evolução do universo exibe *descontinuidades*. O sinequismo consiste em uma 'tendência de considerar tudo como contínuo' (CP 7.565; 1.172, 6.103). Para Peirce (CP 6.169), o sinequismo 'é aquela tendência do pensamento filosófico que insiste na idéia da continuidade como sendo de importância primária na filosofia e, em particular, na necessidade de hipóteses envolvendo verdadeira continuidade.'[11] Defendemos que esta incompatibilidade é apenas aparente, uma vez que a filosofia emergentista pode ser desenvolvida como um modo de superar a dicotomia entre continuidade e descontinuidade. Uma filosofia emergentista dessa natureza pode acomodar o sinequismo de Peirce.

Em seu *Emergent Evolution* (1923), um dos principais teóricos do emergentismo britânico, Conwy Lloyd Morgan, propôs uma teoria da emergência que buscava combinar as idéias de continuidade e

descontinuidade. Entre as teses fundamentais da teoria de Morgan, há duas que têm conseqüências para nossa discussão: as teses da co-ocorrência de emergentes e resultantes, e da continuidade quantitativa e novidade qualitativa. Para Morgan, propriedades emergentes sempre ocorrem acompanhadas de propriedades resultantes, que poderiam ser previstas a partir de conhecimento teórico sobre o nível anterior, e confeririam continuidade ao processo evolutivo. Assim, embora a emergência correspondesse ao surgimento de propriedades genuinamente novas, que não poderiam ser previstas a partir do conhecimento de entidades e processos pré-existentes, ela não corresponderia a um salto no processo evolutivo. Isso é evidente no modo como Morgan entendia a produção de novidade qualitativa na evolução: ela seria uma mudança qualitativa de direção, ou um ponto crítico de mudança em um processo evolutivo contínuo. De acordo com Morgan (1923: 5), '... através dos resultantes, há continuidade no progresso; através da emergência, há progresso na continuidade.'

Considere-se, além disso, que o processo de mudança gradual e quantitativa dos sistemas naturais cria as condições para a mudança qualitativa expressa na noção de emergência, na forma de um ponto crítico de mudança no modo de evolução daqueles sistemas. Tomando a origem da vida como exemplo, Morgan (1923: 7) argumenta tanto a favor de uma 'continuidade resultante entre o não-vivo e o vivo', quanto a favor de uma novidade qualitativa, que para ele não é incompatível com tal continuidade: 'Mas alguém pode ainda perguntar se não há, em algum estágio deste processo, uma nova característica emergente da vida [...]. Parece haver algo genuinamente novo em algum estágio da continuidade resultante.' Em suma, o fato de que uma das primeiras, e mais influentes, teorias da emergência não postule saltos no processo evolutivo sugere que não há contradição entre o sinequismo de Peirce e uma filosofia emergentista.

Podemos especular que a competência para manipular signos tenha aparecido na evolução dos sistemas como produto de um processo contínuo. Entretanto, quando sistemas semióticos aparecem, eles exibem comportamento distinto de sistemas reativos. Sistemas semióticos evoluem de um modo diferente, comparativamente a sistemas reativos; uma mudança qualitativa no modo de evoluir teve lugar com seu surgimento. Um sistema que é capaz de interpretar o mundo, através da mediação de signos, evolui de um

modo que é determinado pelo fato de que é capaz de usar signos para obter informação sobre o ambiente, de tal maneira que os signos cumprem funções que favorecem a sobrevivência e aumentam as chances de sobrevivência e reprodução de seus usuários (Emmeche & El-Hani 2000, Emmeche 1998).

A quarta questão: (4) a semiose pode ser descrita como um processo sistêmico? A atualização de cadeias de tríades potenciais depende de condições de contorno estabelecidas por um nível macro-semiótico. É possível entender o nível macro-semiótico como correspondente ao sistema semiótico como um todo, baseado na idéia de que este último pode ser considerado a *corporificação* (*embodiment*) de processos semióticos (CP 5.314). Embora a semiose seja instanciada no nível focal, ela deve ser entendida como uma propriedade sistêmica, já que o nível macro-semiótico estabelece as condições de contorno requeridas para a sua atualização. Isto é, a própria instanciação da semiose, no nível focal, depende de restrições colocadas pelo sistema semiótico como um todo (nível macro-semiótico).

Quanto à questão 5 (como devem ser descritos níveis em um sistema semiótico?), a seção anterior pode ser considerada uma resposta a ela. Em seguida, perguntamos: (6) em que sentido podemos dizer que a semiose é sincronicamente determinada pelas propriedades e pelo arranjo das partes de um sistema semiótico? Em nosso modelo, a semiose se situa no nível focal, instanciada na forma de cadeias de tríades, enquanto tríades individuais estão situadas no nível imediatamente inferior, e redes de cadeias, no nível imediatamente superior. Assim, ao tratarmos da idéia de determinação sincrônica, temos de concentrar nossa atenção na relação entre cadeias de tríades, no nível focal, e tríades individuais, no nível micro-semiótico.

A semiose é descrita por Peirce como um padrão de relações determinativas entre correlatos especificados funcionalmente. Podemos dizer, então, que a semiose é sincronicamente determinada pela micro-estrutura das tríades individuais que compõem uma cadeia de tríades, i.e., pelas propriedades relacionais e pelo arranjo dos elementos S, O e I.[12] Não pode haver qualquer diferença na semiose sem uma diferença nas propriedades e/ou no arranjo de S, O e I.[13]

As propriedades de S, O e I são relacionais porque estes elementos estão engajados em relações ordenadas triadicamente

irredutíveis. Como Savan (1987-88: 43) afirma, 'os termos interpretante, signo e objeto são uma tríade cuja definição é circular. Cada um dos três é definido conforme os outros dois.' A única propriedade de S, O e I é a maneira como se posicionam, em seus papéis funcionais, uns em relação aos outros, como primeiro, segundo e terceiro termos.

Devemos também considerar a força modal da relação de determinação sincrônica entre cadeias de tríades e tríades. Consideraremos aqui quatro possibilidades (ver Bailey 1999): (i) necessidade fraca, em que a relação de determinação vale no mundo real, mas não vale necessariamente em qualquer outro mundo possível; (ii) necessidade natural ou física, ou nômica, ou nomológica, em que a relação determinativa vale no mundo real e em todos os mundos naturalmente possíveis, que podem ser descritos como todos os mundos nos quais as leis físicas se assemelham àquelas encontradas no mundo real; (iii) necessidade metafísica, em que a relação determinativa vale no mundo real e em todos os mundos metafisicamente possíveis, que abrangem todos os mundos nos quais verdades necessárias *a posteriori* (tais como 'água é H2O') são válidas; (iv) necessidade lógica, em que a relação determinativa vale no mundo real e em todos os mundos logicamente possíveis, incluindo aqueles nos quais verdades necessárias *a priori* se sustentam - este é o conjunto de todos os mundos possíveis.

No caso da semiose, argumentamos, as relações determinativas entre os elementos de tríades individuais, bem como entre tríades, em uma cadeia de tríades, valem com necessidade lógica.[14] Note que a demonstração de que S-O-I constitui uma relação indecomponível deve ser primeiro conduzida logicamente (Queiroz 2004; Houser 1997: 16). A razão da precedência de um tratamento formal de relações sobre um tratamento empírico, e metafísico, reside no fato de que só formalmente pode-se conduzir uma análise das propriedades de completude e suficiência das categorias (Parker 1998: 3, 43). Apenas ulteriormente a propriedade de irredutibilidade lógica é verificada em um domínio empírico e metafísico. A precedência de um tratamento lógico tem importantes conseqüências metodológicas. Uma análise de propriedades formais, em contraste com propriedades materiais[15], deve anteceder qualquer investigação empírica e metafísica das categorias. Em outras palavras, uma análise lógico-matemática das categorias deve ser anterior a qualquer formulação nos âmbitos da fenomenologia, das ciências normativas e da

metafísica[16], que empregam técnicas e resultados matemáticos para validar as categorias (Hookway 1985: 182).

Portanto, em nossa discussão sobre a força modal da relação de determinação sincrônica entre tríades e cadeias de tríades, começamos com um tratamento lógico das relações entre os elementos da semiose. Focamos nossa atenção, primeiro, nos papéis funcionais de S, O e I, conforme estabelecidos em uma análise lógica de relações. Os papéis funcionais de S, O e I são logicamente determinados em cada tríade, no que diz respeito tanto às relações internas à tríade quanto à constituição de cadeias de tríades. Estas relações determinativas valem com *necessidade lógica*: em um mundo substancialmente diferente do mundo real em suas leis físicas, i.e. em um mundo nomologicamente distinto do mundo real, as relações lógicas entre S, O e I seriam as mesmas.

Se estivermos certos, as relações determinativas entre S, O e I são logicamente válidas no conjunto de todos os mundos possíveis, desde que o mundo concebido admita a existência de entidades ou processos materiais. Afinal, há uma importante restrição para que alguma coisa seja um sistema semiótico — ela deve ser materialmente incorporada. Isso não implica que as relações determinativas entre S, O e I poderiam ser apenas nomologicamente válidas, mas que qualquer mundo logicamente concebível, no qual a semiose possa ter lugar, é um mundo cujas leis permitem a existência de entidades e processos materiais, que são uma condição necessária para semiose. Em tal mundo, as relações determinativas entre S, O e I valem com necessidade lógica. Se supusermos que existam mundos logicamente concebíveis onde nenhuma matéria esteja presente, isso implicaria que tais mundos não exibiriam qualquer sistema ou processo semiótico, e nenhuma relação de determinação entre S-O-I teria lugar nestes mundos.

Em um domínio empírico, devemos focar nossa atenção não apenas nos papéis funcionais de S, O e I, mas também no modo como tais papéis *podem* ser incorporados (*embodied*) e no modo como as relações entre S, O e I podem ser instanciadas no mundo real. Neste caso, deve-se notar que, enquanto os papéis funcionais são logicamente determinados, os ocupantes dos papéis funcionais de S, O e I são contingentes. Que a palavra 'elefante', por exemplo, seja um signo de um grande animal no mundo é usualmente tratado como algo contingente. Não é logicamente necessário que a palavra 'elefante', S, esteja, por meio de I, para um grande ani-

mal, O. Mas as relações determinativas entre estes elementos são logicamente determinadas e os papéis funcionais de S, O e I também o são. Assim, em um mundo suficientemente distinto do mundo real, em suas leis físicas, entidades ou processos inteiramente diferentes poderiam ocupar os papéis funcionais de S, O e I, em distintos sistemas semióticos. Podemos concluir que o fato de que uma certa classe de entidades, ou processos, atue funcionalmente em um processo semiótico vale com grau de necessidade nomológica, e não lógica, ainda que o papel funcional valha com grau de necessidade lógica.

A questão seguinte (7) se refere ao modo como devemos entender o princípio de irredutibilidade da semiose. A relação semiótica triádica é descrita por Peirce como *irredutível* no sentido de que não pode ser decomposta em relações mais simples:

> Por semiose, eu quero dizer [...] uma ação, ou influência, que é, ou envolve, a cooperação de três sujeitos, tais como um signo, seu objeto, e seu interpretante, esta influência tri-relativa não podendo, de modo algum, ser resolvida em termos de ações entre pares (CP 5.484).

Como Peirce discute cuidadosamente a irredutibilidade de tríades, vamos considerar o que definimos acima como nível micro-semiótico. Primeiro, a relação semiótica *não* é irredutível porque a condição de analisabilidade é violada. As propriedades que uma tríade possui, incluindo a propriedade de ser semiótica, segue do comportamento dos elementos da tríade. Se conhecermos as relações de quaisquer três elementos, então seremos capazes de saber se o processo em que estão envolvidos é semiótico, uma vez que saberemos se os elementos se posicionam lógico-funcionalmente como S, O e I. Dito de outra maneira, a não-analisabilidade não é a razão pela qual deveríamos considerar a semiose como irredutível.

Podemos entender por que uma relação semiótica é irredutível com base na segunda noção de irredutibilidade discutida acima, baseada na não-dedutibilidade do comportamento das partes de um sistema. Neste caso, nós deveríamos mostrar que o comportamento específico dos elementos de uma tríade não segue do comportamento dos elementos em relações mais simples. Notem que a semiose pode ser considerada o melhor exemplo de uma relação triádica na qual a segunda condição de redutibilidade é violada, já que o comportamento dos elementos de uma relação semiótica não segue do comportamento que eles apresentam em isolamento, ou em relações mais simples (diádicas).

Os papéis funcionais dos elementos não podem ser identificados em estruturas mais simples do que em uma relação triádica. O papel funcional de S só pode ser identificado na *relação de mediação* que ele estabelece entre O e I. Similarmente, o papel funcional de O é identificado na relação em que ele determina I por meio de S. E o papel de I é identificado pelo fato de que ele é determinado por O através de S. Se considerarmos apenas relações diádicas, S-I, S-O, ou I-O, ou se considerarmos os elementos em isolamento, não poderemos inferir o comportamento de tais elementos em uma relação triádica S-O-I (ver EP 2:391). A irredutibilidade da semiose deve ser entendida em termos da não-dedutibilidade do comportamento dos elementos lógico-funcionais de uma tríade, a partir de seus comportamentos em relações mais simples.

É importante lembrar, contudo, que a realização (*actualization*) de uma tríade individual depende de restrições estabelecidas por um nível macro-semiótico, que seleciona, entre um conjunto de tríades potenciais, aquelas tríades que serão instanciadas para formar cadeias de tríades. Argumentamos que é somente no contexto de tais cadeias que tríades individuais podem ser definidas. Embora seja o caso que a tese da irredutibilidade da semiose esteja apoiada na natureza das relações entre S, O e I, é também o caso que, para serem atualizadas, estas relações dependem de restrições estabelecidas pelos níveis macro-semiótico e focal. Assim, a semiose é um processo irredutível, emergente, do sistema semiótico, e não de uma tríade qualquer.

Quanto à oitava questão, as estruturas de tríades e cadeias de tríades podem ser consideradas imprevisíveis, já que Peirce defende o indeterminismo e o acaso como fatores fundamentais no universo. Assim, o comportamento dos elementos em um processo semiótico é também imprevisível a partir dos comportamentos que podem exibir em sistemas mais simples. Pode-se afirmar que a semiose é um processo emergente, que apresenta uma estrutura imprevisível em princípio, em virtude da natureza indeterminística do processo evolutivo. Este argumento está baseado na tese peirceana do tiquismo, que consiste na defesa metafísica do 'acaso absoluto' como um fator real no universo (ver Murphey 1993, Potter 1997). O tiquismo tem um papel essencial na cosmologia evolutiva de Peirce, sendo considerado por ele a única explicação para a multiplicidade e irregularidade encontradas no universo.

O ponto mais importante aqui é que, de acordo com a cosmologia evolutiva de Peirce, tudo deveria ser explicado como um produto de processos evolutivos que têm estados de indeterminação e acaso como ponto de partida. Pape (2002: 226), em um artigo sobre o tiquismo e processos mentais, afirma que 'a matéria, o tempo, o espaço e as próprias leis da natureza — todas elas devem ser explicadas como regularidades de interação emergentes que surgem de um estado de indeterminação'. Isso sugere que o emergentismo é compatível com doutrinas centrais da metafísica de Peirce, como o tiquismo e o sinequismo.

Os argumentos desenvolvidos nesta seção permitem-nos concluir que, no caso da compreensão dos processos semióticos numa moldura teórica peirceana, uma teoria forte da emergência pode ser defendida. Esta teoria deve incluir: (1) um conceito de irredutibilidade baseado na não-dedutibilidade do comportamento de signos, objetos e interpretantes em relações triádicas, a partir de seus possíveis comportamentos em relações mais simples; e (2) uma tese da imprevisibilidade em princípio da estrutura dos processos semióticos.

Conclusão

De acordo com Rosenthal (1994: 27), 'significados devem ser entendidos como estruturas relacionais que emergem de padrões de comportamento.' Uma avaliação precisa desta afirmação depende de uma compreensão clara sobre como 'estruturas' constituem propriedades sistêmicas emergentes, e sobre o modo como se relacionam à micro-estrutura de uma certa classe de sistemas. Este é um exemplo típico de uma situação em que não é conveniente usar a idéia de emergência de um modo ordinário. Mencionamos, no começo deste capítulo, outra situação, surgida no contexto de nosso próprio trabalho sobre simulações computacionais de processos semióticos, em que é necessário caracterizar a semiose como uma propriedade ou um processo emergente de maneira precisa (e.g. Gomes et al., prelo). Não encontramos, contudo, um tratamento deste problema em periódicos e livros dedicados à semiótica. Esta foi uma das motivações para discutirmos, neste trabalho, as condições que devem ser satisfeitas para que a semiose possa ser caracterizada como um processo emergente.

O procedimento que empregamos consistiu no levantamento de questões que devem ser respondidas para que o conceito de emergência seja usado de modo preciso, no domínio dos fenômenos semióticos. A lista de questões que formulamos, e para as quais oferecemos respostas na seção anterior, é a seguinte: (1) O que é um sistema semiótico? (2) Os sistemas que exibem semiose são constituídos apenas fisicamente? (3) Os sistemas semióticos constituem uma nova classe de sistemas, instanciando novas estruturas, processos, propriedades, disposições etc.? (4) A semiose pode ser descrita como um processo sistêmico? (5) Como devem ser descritos níveis em um sistema semiótico? (6) Em que sentido podemos dizer que a semiose, como um processo emergente em sistemas semióticos, é determinada sincronicamente pelas propriedades e pelo arranjo das partes do sistema? (7) Qual interpretação da irredutibilidade é mais adequada para dar conta da idéia de irredutibilidade da semiose? (8) A estrutura de sistemas ou processos semióticos pode ser considerada imprevisível em princípio? (9) Alguma forma de causação descendente poderia estar envolvida na semiose?

Concluímos que uma teoria forte da emergência pode ser defendida no caso dos processos semióticos, incluindo um conceito de irredutibilidade baseado na não-dedutibilidade do comportamento de signos, objetos e interpretantes em relações triádicas, e em uma tese da imprevisibilidade em princípio da estrutura dos processos semióticos.

Utilizamos o estruturalismo hierárquico de Salthe como base para a proposição de um modelo para explicar emergência de semiose em sistemas semióticos. Conforme este modelo, a semiose é entendida como um processo sistêmico no nível focal, em que cadeias de tríades são instanciadas em decorrência da interação entre potencialidades estabelecidas em um nível micro-semiótico (condições iniciadoras) e a influência seletiva, regulatória, de um nível macro-semiótico (condições de contorno).

Nossa expectativa é a de que as questões que propusemos para uma caracterização precisa da semiose como um processo emergente, e a modelagem deste processo em um modelo incluindo três níveis, baseado no estruturalismo hierárquico de Salthe, contribuam para um diálogo consistente entre os pensamentos emergentista e semiótico.

AGRADECIMENTOS

C.N.E.H. e J.Q. agradecem ao CNPq e à FAPESB.

Notas

[1] Parte deste trabalho foi publicado em *Galaxia* 9: 113-140, 2005.

[2] A segunda maneira de explicar a origem da ordem pode ser também entendida como uma explicação da produção da ordem a partir do caos, se atribuirmos ao termo 'caos' o sentido que ele tinha na filosofia Grega clássica, *i.e.*, o de ausência de forma.

[3] Cariani (1989: 98) define um aparato formal-computacional como um aparato no qual o comportamento de transição de estados observado ao longo de um período observacional pode ser completamente descrito em termos de computações. A descrição de um comportamento como uma computação requer que (1) um quadro observacional fixo seja especificado, podendo ser comunicado a outro observador de modo que este replique as observações; (2) para cada estado observado, não pode haver mais que um estado sucessor imediato, de modo que regras determinísticas de transição de estados possam ser construídas para todos os estados; e (3) um estado final é alcançado após uma extensão de tempo finita (Cariani 1989, cap. 5). O conceito de computação é intimamente relacionado ao conceito de 'procedimento formal', que se caracteriza pela enorme confiabilidade com a qual leva aos mesmos resultados. Para Cariani (1989: 99-100), um procedimento formal é (1) determinístico; (2) executável em tempo finito por meio de um equipamento finito; (3) de execução 'mecânica' ou 'construtiva', passível de descrição precisa, de modo que outra inteligência, ou talvez outro aparato, ao receber a descrição, seja capaz de aplicar o procedimento e obter resultados idênticos; e (4) passível de representação em termos numéricos.

[4] Neste ponto, surge uma tensão entre a crítica feita por Cariani à emergência em simulações computacionais e o conceito (epistêmico) de emergência relativa a um modelo que ele emprega para compreender a emergência em sistemas biológicos; afinal, esta última forma de emergência também é dependente do observador.

[5] Outra dificuldade com a concepção computacionalista da emergência diz respeito a um velho problema enfrentado pelo emergentismo, a 'ameaça da trivialização'. A emergência computacional parece absolutamente ubíqua, na medida em que, se definimos um comportamento ou propriedade como 'emergente' com base em nossa incapacidade de prever o resultado de uma

computação, praticamente todas as computações se tornam processos de emergência (Cariani 1989: 149).

[6] Contudo, a emergência combinatorial também ocorre, de acordo com Cariani (1997), na evolução biológica, por exemplo, em eventos nos quais novas seqüências de nucleotídeos em uma molécula de DNA surgem a partir de combinações de seqüências preexistentes, como em eventos de 'mistura de exons' (*exon shuffling*), entre outros casos.

[7] Mais recentemente, Boogerd (et al. 2005) apresentaram uma visão mais elaborada das noções de irredutibilidade como 'não-analisabilidade' e 'não-dedutibilidade'.

[8] Notem que as duas razões para a imprevisibilidade das propriedades emergentes têm *status* diferentes. Enquanto a segunda é de natureza empírica, particularmente se a irredutibilidade for interpretada em termos da não-dedutibilidade, a primeira depende de um 'compromisso metafísico'.

[9] Há uma clara correspondência entre a estrutura hierárquica proposta por Salthe e a distribuição hierárquica das categorias de Peirce. Níveis micro-semióticos devem garantir que processos sígnicos apresentem comportamentos de indeterminação e, neste nível, *iniciam-se* os processos semióticos. A associação com a Primeiridade é direta. Em níveis focais, os processos são espaço-temporalmente instanciados, produzindo *tokens*, que são exemplos de Secundidade. Níveis macro-semióticos, no domínio da Terceiridade, garantem *generalidade* e *temporalidade* aos processos sígnicos, que se tornam histórico e contexto-dependentes.

[10] A regulação de um processo no nível focal por condições de contorno estabelecidas pelo nível superior é entendida aqui como um tipo de processo seletivo. Suponha que uma relação causal entre um dado elemento de um sistema, *A*, e outro elemento do mesmo sistema, *B*, seja regulada. Essa relação é entendida como a seleção de *B* como o efeito de *A*, entre uma diversidade de efeitos possíveis, pelas condições de contorno estabelecidas por estruturas de nível superior, nas quais a relação causal em questão está inserida. Esta idéia se baseia nas contribuições de Polanyi (1968) e Campbell (1974), e está relacionada ao problema da causação descendente.

[11] Para uma abordagem detalhada do sinequismo, ver Parker (1998), Potter (1997), Murphey (1993).

[12] Para uma compreensão precisa de nosso argumento, é importante não confundir determinação sincrônica e diacrônica. Defendemos que o quadro teórico de Peirce acomoda a tese de uma determinação sincrônica, enquanto claramente rejeita uma determinação diacrônica.

[13] O arranjo dos elementos S-O-I é especificado pelas relações de determinação entre eles. De outro modo, a tríade seria uma mera justaposição de três elementos (CP 1.371, 1.363; Brunning 1997, De Tienne 1992).

[14] Observe que, no âmbito das discussões sobre as relações lógicas entre elementos e tríades, trabalhamos no domínio da Gramática Especulativa (ver CP 1.444). Para Houser (1997: 9), 'o lógico que se concentra na gramática especulativa investiga as relações de representação (signos), procura elaborar as condições necessárias e suficientes para a representação, e classifica os diferentes tipos possíveis de representação'.

[15] A divisão entre propriedades materiais e formais das categorias foi claramente estabelecida por Peirce depois de 1885 (Kent 1997: 448).

[16] Para uma introdução à fenomenologia, às ciências normativas e à metafísica, ver De Waal (2001), Parker (1998).

Referências

BAAS, N. A. 1996. A framework for higher-order cognition and consciousness. Em: Toward a Science of Consciousness. Hameroff, S.R.; Kaszniak, A. W.; Scott, A. C. (eds.). MIT Press. pp. 633-648.

BAAS, N. A.; EMMECHE, C. 1997. On emergence and explanation. *Intellectica* 25: 67-83.

BAILEY, A. 1999. Supervenience and physicalism. *Synthese* 117: 53-73.

BEDAU, M. 2002. Downward causation and autonomy of weak emergence. *Principia 6* (1): 5-50.

BERGMAN, M. 2000. Reflections on the role of the communicative sign in semeiotic. *Transactions of the Charles S. Peirce Society:* A Quarterly Journal in American Philosophy, Spring XXXVI (2): 225-254.

BLITZ, D. 1992. *Emergent Evolution:* Qualitative Novelty and the Levels of Reality. Dordrecht: Kluwer.

BOOGERD, F. C.; BRUGGEMAN, F. J.; RICHARDSON, R. C.; STEPHAN, A. & WESTERHOFF, H. 2005. Emergence and its place in nature: a case study of biochemical networks. *Synthese* 145(1): 131-164.

BRUNNING, J. 1997. Genuine triads and teridentity. Em: S*tudies in the logic of Charles Sanders Peirce*. N. Houser, D. Roberts, J. Evra (eds.). Indiana University Press. pp. 252-270.

BURCH, R. 1991. *A Peircean Reduction Thesis*. Texas Tech University Press.

__. 1997. Peirce's reduction thesis. Em: *Studies in the logic of Charles Sanders Peirce*. N. Houser, D. Roberts, J. Evra (eds.) Indiana University Press. pp. 234-251.

CAMPBELL, D. T. 1974. Downward causation in hierarchically organised biological systems. Em: *Studies in the Philosophy of Biology:* Reduction and Related Problems. F. Ayala e Th. Dobzhansky (eds.). University of California Press. pp. 179-186.

CARIANI, P. 1989. *On the Design of Devices with Emergent Semantic Functions*, PhD Dissertation, Department of Systems Science, State University of New York, Binghamton.

__. 1991. Emergence and artificial life, Artificial Life II. Em: *SFI Studies in the Sciences of Complexity,* Proc. Vol. X. C. C. Langton, J.D. Taylor, S. Farmer, Rasmussen (eds). Addison-Wesley. pp.775-797.

__. 1997. Emergence of new signal-primitives in neural systems. *Intellectica 2:* 95-143.

CLARK, A. 1997. *Being There:* Putting Brain, Body and World Together Again. MIT Press, Bradford Books.

__. 2001. *Mindware:* An Introduction to the Philosophy of Cognitive Science. Oxford University Press.

DEACON, T. 1999. Memes as signs. *The Semiotic Review of Books* 10(3): 1-3.

DE TIENNE, A. 1992. Peirce's semiotic monism. Em: *Signs of Humanity/L'homme et ses signes*. Proceedings of the IVth International Congress/Actes du IVe Congrès Mondial. G. Deledalle, M. Balat e J. Rhodes Deledalle. (eds.). (Approaches to Semiotics. 107), Mouton de Gruyter. pp. 1291-1303.

__. 2003. Learning *qua* semiosis. *S.E.E.D. Journal -- Semiotics, Evolution, Energy, and Development* (3): 37-53.

DE WAAL, C. 2001. *On Peirce*. Thomson Learning.

EMMECHE, C. 1994. *The Garden in the Machine:* The Emerging Science of Artificial Life. Princeton University Press.

__. 1997. *Defining life, explaining emergence*. http://www.nbi.dk/~emmeche/

__. 2003. Causal processes, semiosis, and consciousness. Em: *Process Theories:* Crossdisciplinary Studies in Dynamic Categories. J. Seibt (ed.). Kluwer. pp. 313-336.

EMMECHE, C.; KØPPE, S. & STJERNFELT, F. 1997. *Explaining emergence:* towards an Ontology of Levels. Journal for General Philosophy of Science 28: 83-119.

EL-HANI, C. N. 2002. On the reality of emergents. *Principia 6*(1): 51-87.

EL-HANI, C.N., EMMECHE, C. 2000. *On some theoretical grounds for an organism-centered biology:* Property emergence, supervenience, and downward causation. Theory in Biosciences 119: 234-275.

EL-HANI, C. N.; QUEIROZ, J. & EMMECHE, C. 2006. A semiotic analysis of the genetic information system. *Semiotica* 160 (1/4): 1-68.

FETZER, J. H. 1997. Thinking and computing: computers as special kinds of signs. *Minds and Machines* 7: 345-364.

__. 1988. Signs and minds: an introduction to the theory of semiotic systems. Em: *Aspects of Artificial Intelligence*. J. Fetzer (ed). Dordrecht. pp. 133-161.

FISCH, M. 1986. *Peirce, Semeiotic, and Pragmatism*. Indiana University Press.

GOMES, A.; GUDWIN, R.; QUEIROZ, J. 2003a. On a computational model of Peircean semiosis. Em: *Proceedings of the International Conference on Integration of Knowledge Intensive Multi-Agent Systems - KIMAS'03*, IEEE. H Hexmoor (ed.). pp. 703-708.

__. 2003b. Towards meaning processes in computers from Peircean semiotics. *S.E.E.D. Journal -- Semiotics, Evolution, Energy, and Development* 3 (2): 69-79.

GOMES, A.; EL-HANI, C. N.; GUDWIN, R.; QUEIROZ, J. prelo. Towards the emergence of meaning processes in computers from Peircean

semiotics. *Mind & Society - Cognitive Studies in Economics and Social Sciences*.

HAUSMAN, C. 1993. *Charles Sanders Peirce's Evolutionary Philosophy*. Cambridge University Press.

HOOKWAY, C. 1985. *Peirce*. Routledge & Kegan Paul.

HOUSER, N. 1997. Introduction: Peirce as a logician. Em: *Studies in the logic of Charles Sanders Peirce*. N. Houser, D. Roberts, J. Evra (eds.) Indiana University Press. pp. 1-22.

KENT, B. 1997. The interconnectedness of Peirce's diagrammatic thought. Em: *Studies in the Logic of Charles S. Peirce*. N. Houser, D. Roberts, J. Evra (eds.). Indiana University Press. pp. 445-459.

KETNER, K . 1986. Peirce's most lucid and interesting paper: an introduction to cenopythagoreanism. *International Philosophical Quarterly* 26: 375-392.

KIM, J. 1998. *Mind in a Physical World:* An Essay on the Mind-Body Problem and Mental Causation. MIT Press.

__. 1999. Making sense of emergence. *Philosophical Studies* 95: 3-36.

KLEE, R. L. 1984. Micro-determinism and concepts of emergence. *Philosophy of Science* 51 (1), 44-63.

LANGTON, C. 1989. Artificial life. Em: *Artificial Life, Santa Fe Institute Studies in the Sciences of Complexity*. Vol. 5. (Proceedings of the First Conference on Artificial Life, Los Alamos, September, 1987). C. Langton (ed.) Addison-Wesley.

MCLAUGHLIN, B. P. 1992. The rise and fall of British emergentism. Em: *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. A. Beckermann, H. Flohr e J. Kim (eds.). Walter de Gruyter. pp. 49-93.

MORGAN, C. L. 1923. *Emergent Evolution*. Williams and Norgate. Disponível em http://spartan.ac.brocku.ca/~lward/morgan/Morgan_1923

MERRELL, F. 1995. *Peirce's Semiotics Now*. Canadian Scholar's Press.

MURPHEY, M. G. 1993. *The Development of Peirce's Philosophy*. Hackett.

NAGEL, E. 1961. *The Structure of Science*. Harcourt, Brace, and World.

PAPE, H. 2002. What thought is for: the problematic identity of mental processes with chance events in Peirce's idealistic metaphysics. *Transactions of the Charles S. Peirce Society:* A Quarterly Journal in American Philosophy, Winter/Spring, XXXVIII (1/2): 215-251.

PARKER, K. 1998. *The Continuity of Peirce's Thought*. Vanderbilt University Press.

PEIRCE, Charles S. 1992 e 1998 (EP1, EP2). *The Essential Peirce. Selected Philosophical Writings*. vol. 1, 1867-1893, N. Houser e C. Kloesel (eds.). vol. 2, 1893-1913, the Peirce Edition Project (ed.). Indiana University Press.

PEIRCE, C.S. 1931-1935. *The Collected Papers of Charles Sanders Peirce*. Electronic edition reproducing vols. I-VI, Hartshorne, C. e Weiss, P. (eds.), Cambridge: Harvard University Press, 1931-1935; vols. VII-VIII A. W. BURKS (Ed.), same publisher, 1958. Charlottesville: Intelex Corporation. (citado como CP, seguido por volume e número do parágrafo.)

PIHLSTRÖM, S. 2002. *The re-emergence of the emergence debate*. Principia 6: 133-181.

POLANYI, M.1968). *Life's irreducible structure*. Science 160: 1308-1312.

POPPER, K.R. & ECCLES, J.C. [1977] 1986. *The Self and Its Brain*. Routledge and Kegan Paul.

POTTER, V. 1997. *Charles S. Peirce:* On Norms & Ideals. University of Massachusetts Press.

QUEIROZ, J. 2004. *Semiose Segundo C. S. Peirce*. Educ/Fapesp.

QUEIROZ, J. & EL-HANI, C. N. 2006a. Semiosis as an emergent process. *Transactions of the Charles Sanders Peirce Society* 42 (1): 78-116.

__. 2006b. Towards a multi-level approach to the emergence of meaning processes in living systems. *Acta Biotheoretica* 54 (3):174-206.

RANSDELL, J. 1977. Some leading ideas of Peirce's semiotic. *Semiotica* 19 (3/4): 157-178.

RONALD, E.M.A.; SIPPER, M.; CAPCARRÈRE, M.S. 1999. Design, observation, surprise! A test for emergence. *Artificial Life* 5: 225-239.

ROSENTHAL, S. 1994. *Charles Peirce's Pragmatic Pluralism*. State University of New York Press.

SALTHE, S. N. 1985. *Evolving Hierarchical Systems:* Their Structure and Representation. Columbia University Press.

SANTAELLA, L. 1995. *A Teoria Geral dos Signos:* semiose e autogeração. Editora Ática.

SAVAN, D. 1986. Response to T.L.Short. *Transactions of the Charles S. Peirce Society:* A Quarterly Journal in American Philosophy, Summer XXII (2): 125-143.

__. 1987-1988. *An Introduction to C. S. Peirce's Full System of Semeiotic*. Toronto Semiotic Circle, Monograph Series of the TSC, Number 1.

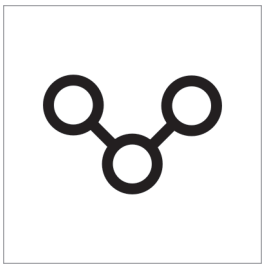
STEPHAN, A. 1998. Varieties of Emergence in Artificial and Natural Systems. *Zeitschrift für Naturforschung* 53c: 639-656.

STEPHAN, A. 1999. *Emergenz:* Von der Unvorhersagbarkeit zur Selbstorganisation. Dresden and München: Dresden University Press.

COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA



CAPÍTULO 4

O QUE É O SÍMBOLO

Lucia Santaella

No campo das mais diferenciadas ciências e artes, a palavra símbolo foi e continua sendo empregada com tal generosidade que seu sentido se envolveu em brumas. A definição peirceana, ao contrário, é técnica e precisa. Para chegar a ela, devemos começar pelo entendimento do legi-signo, pois é nele que o símbolo encontra seu suporte.

Legi-signo é uma lei que é um signo. Antes de tudo, é preciso considerar que a noção peirceana de lei é muito original (ver Santaella 1999a,b). Lei não se confunde com necessidade, nem estritamente com norma, pois esta é apenas uma tradução convencional da lei. Para Peirce, a lei é uma força viva, uma 'força condicional permanente' (CP 3.435), quer dizer, é uma 'regularidade no futuro indefinido' (CP 2.293). Sem o governo da lei, fatos e ações são brutos e cegos. Conformando-se, até certo ponto, à força viva da lei, os fatos se acomodam dentro de uma regularidade, de certo modo, previsível. A lei funciona, portanto, como uma força que será atualizada, dadas certas condições. Por isso mesmo, a lei não tem a

rigidez de uma necessidade, podendo ela própria evoluir, transformando-se. Contudo, em si mesma, a lei é uma abstração. Ela não tem existência concreta a não ser através dos casos que governa, casos que nunca poderão exaurir todo o potencial de uma lei como força viva. Quer dizer, a lei que governa os fatos é geral, enquanto os fatos são particulares, mas ao mesmo tempo, a lei lhes empresta uma certa generalidade que se expressa através da regularidade.

Tendo isso em vista, 'o legi-signo é um signo considerado no que diz respeito a um poder que lhe é próprio de agir semioticamente, isto é, de gerar signos interpretantes' (Ransdell 1983: 54). A lei de representação já está contida no próprio signo, de modo que ele está fadado a produzir um signo interpretante ou uma série de signos interpretantes tão gerais quanto ele próprio, através dos quais seu caráter de signo se realiza. É a lei que fará o signo ser interpretado como sendo um signo, pois o legi-signo funciona como uma regra que irá determinar seu interpretante, uma regra que determinará que ele seja interpretado como se referindo a um dado objeto.

A linguagem verbal é o exemplo mais evidente de legi-signo ou sistema de legi-signos. Por pertencerem ao sistema de uma língua, as palavras são interpretadas como representando aquilo que representam por força das leis desse sistema. Como quaisquer outros exemplares de legi-signo, no seu estatuto de leis, as palavras só tomam parte na experiência ou têm existência concreta por meio de suas manifestações. Peirce chama de 'réplicas' essas instâncias de manifestação. Tratam-se de sin-signos de tipo especial. São sin-signos porque são existentes individuais que ocorrem em um tempo e espaço determinado, mas são réplicas porque atualizam, corporificam legi-signos. 'O legi-signo é uma classe das réplicas da palavra, mas nenhuma coleção finita de réplicas poderá exaurir a classe. Mesmo assim, a existência do legi-signo está nos enunciados e inscrições individuais de suas réplicas' (Savan 1976: 29), conforme a passagem abaixo pode melhor esclarecer.

> Falamos de escrever ou pronunciar a palavra 'homem' , mas isso é apenas uma réplica ou materialização da palavra que é pronunciada ou escrita. A palavra, em si mesma, não tem existência, embora tenha ser real, consistindo em que os existentes deverão se conformar a ela. É um tipo geral de sucessão de sons, ou *representamens* de sons, que só se torna um signo pela circunstância de que um hábito ou lei adquirida

> levam as réplicas, a que essa sucessão dá lugar, a serem interpretadas como significando um homem. Tanto as palavras quanto seus signos são regras gerais, mas a palavra isolada determina as qualidades de suas próprias réplicas (CP 2.292).

O que vale para as palavras, vale do mesmo modo para as expressões lingüísticas e para os padrões de frases que também se constituem em tipos gerais abstratos. Ora, o tipo geral é a lei que fará as réplicas se conformarem a ela. Por mais variações qualitativas que possam existir nas manifestações concretas, nas réplicas orais ou escritas de uma palavra ou de um padrão frasal, elas sempre se conformarão a uma invariância que é a da palavra ou do padrão como lei. Por isso mesmo, a essência de um legi-signo é formal e não material. A materialidade lhe é emprestada pelos sin-signos nos quais se corporifica, ao mesmo tempo, que, como lei, empresta a eles generalidade. As conseqüências do caráter formal e não apenas material da lei estão expressas na passagem a seguir.

> Suponhamos que eu apague esta palavra 'seis' e escreva 'Seis'. Não se tem aí uma segunda palavra, mas sim, a primeira novamente. Elas são idênticas. Ora, pode a identidade ser interrompida ou devemos dizer que a palavra existia, embora não estivesse escrita? Esta palavra 'seis' implica que duas vezes três é cinco mais um. Esta é uma verdade eterna, a verdade que sempre é e será verdade; e que seria verdade, embora não houvesse no universo seis coisas que pudessem ser contadas, dado que ainda seria verdadeiro que cinco mais um teriam sido duas vezes três. Ora, essa verdade é a palavra SEIS; se por seis entendemos não este traço de giz, mas aquilo em que concordam seis, six, sex, sechs, zes, sei (CP 7.593).

É por isso também que podemos escrever a palavra 'estrela', por exemplo, mas isso não nos faz criadores dessa ou de qualquer palavra. Se apagarmos o que escrevemos, a palavra não terá sido destruída — 'O vocábulo continuará vivendo no espírito daqueles que o empregam. Ainda que todos estejam adormecidos, existe em suas memórias' (CP 2.301). E mesmo que a palavra não esteja mais viva, como é o caso das línguas mortas, nem assim ela perderá seu poder de denotar e significar, pois esse poder lhe é dado por seu caráter de lei, num sistema de leis de que ela é parte indissociável. Dessas características do legi-signo, decorre a natureza do símbolo.

O significado que Peirce deu ao termo 'símbolo', o de 'um signo convencional que depende de um hábito inato ou adquirido' (CP 2.297), não é novo, pois corresponde a um retorno ao seu significado original. Em grego, significava celebração de um contrato ou

convenção. Em Aristóteles, um nome é um símbolo, signo convencional. Os gregos também consideravam como símbolos 'uma fogueira como sinal combinado, um estandarte ou insígnia, uma senha, um emblema, um credo religioso quando serve como distintivo ou traço característico; eram ainda símbolos uma entrada de teatro ou qualquer bilhete ou documento que dá a alguém o direito de receber alguma coisa' (Santaella e Nöth 1998: 63).

Símbolos são signos que funcionam como tal 'não em virtude de um caráter que lhes pertence como coisas, nem em virtude de uma conexão real com seus objetos, mas simplesmente em virtude de serem representados como sendo signos' (CP 8.119). Diferentemente tanto do ícone, que tem sua relação com um possível objeto fundada em uma mera semelhança, quanto do índice, cuja relação com o objeto é uma relação de fato, existencial, o fundamento da relação do símbolo com o objeto que ele representa depende de um caráter imputado, arbitrário, não motivado. Assim, o símbolo é um signo que se conecta 'com seu objeto por meio de uma convenção de que ele será assim entendido, ou ainda por meio de um instinto ou ato intelectual que o toma como representando seu objeto, sem que qualquer ação necessariamente ocorra para estabelecer uma conexão factual entre signo e objeto' (CP 2.308).

O símbolo em si mesmo, na sua natureza de legi-signo, é um tipo geral, abstrato. Não menos abstrato do que o símbolo é seu objeto. Por exemplo, qual é o objeto do legi-signo 'homem'?

> [...] O legi-signo se refere a todos os homens que poderiam logicamente existir - à espécie humana. 'Homem' é um signo coletivo e seu objeto é um necessitante. A pergunta de Peirce é: como pode um legi-signo coletivo, 'homem', ser posto em relação de signo-objeto com a classe geral dos homens? Sua resposta é a de que a palavra deve ser interpretada como sendo o signo de seu objeto. Apenas por meio do interpretante, uma palavra pode ser um signo de uma classe ou uma lei. O símbolo, portanto, é esse signo que se relaciona com seu objeto pelo seu interpretante (Savan 1976: 29).

Portanto, o objeto do símbolo não é algo particular, mas um tipo de coisa, que corresponde a uma idéia ou lei geral a que o símbolo, também como lei, está associado através de uma regra ou hábito interpretativo que Peirce chamava de interpretante lógico. Conclusão: não só o símbolo, mas também seu objeto e ainda seu interpretante são todos os três de natureza geral, tipos abstratos. Vem daí o poder auto-reprodutor do símbolo, pois ele só se consti-

tui como tal através do interpretante (NEM 4:260), conforme está expresso na passagem a seguir.

> O signo é uma relação conjunta com a coisa denotada e com a mente. Se essa relação tripla não é de uma espécie degenerada, o signo se relaciona com seu objeto apenas em conseqüência de uma associação mental, e depende de um hábito. Tais signos são sempre abstratos e gerais, porque hábitos são regras gerais às quais o organismo se submeteu. Na maior parte das vezes, eles são convencionais e arbitrários, incluindo as palavras gerais, o corpo principal da fala, ou qualquer outro modo de se transmitir um julgamento. Por razões de brevidade, eu os chamarei de *tokens* (CP 3.360).

Portanto, 'o símbolo está conectado a seu objeto em virtude de uma idéia da mente que usa o símbolo, sem o que uma tal conexão não existiria' (CP 2.299). Isso significa que 'o símbolo perderia o caráter que faz dele um signo, se não houvesse um interpretante' (CP 2.304). Implícito nessas citações está o fato de que o símbolo é social por natureza, dependendo do uso que uma comunidade faz dele. Conseqüentemente, o terceiro membro da tríade, o interpretante, também se constitui em um tipo geral, trans-individual, ele igualmente uma lei: 'o símbolo é um signo que se refere ao objeto que ele denota em virtude de uma lei, usualmente uma associação de idéias que opera de modo a fazer com que o símbolo seja interpretado como se referindo àquele objeto' (CP 2.249). Ou ainda: 'O valor significativo de um símbolo consiste em uma regularidade associativa, de modo que a identidade do símbolo repousa nessa regularidade' (CP 4.500). A partir de 1906, Peirce passou a chamar essa lei ou regularidade de interpretante lógico, uma regra interpretativa que guia a associação de idéias ligando o símbolo ao seu objeto.

Em muitas passagens, Peirce pôs ênfase no caráter habitual da associação de idéias em virtude da qual o símbolo denota seu objeto, como se pode constatar na seqüência de citações abaixo selecionada.

> [Símbolos] denotam seus objetos apenas em virtude de haver um hábito que associa sua significação com eles (CP 4.544).

> [O símbolo] será interpretado como denotando seu objeto em conseqüência de um hábito (termo que uso inclusive para uma disposição natural). [...] Um símbolo incorpora um hábito e é indispensável para a aplicação de qualquer hábito intelectual, pelo menos. [...] Os símbolos repousam exclusivamente em hábitos já definitivamente formados (CP 4.531).

> Defino um símbolo como um signo que é determinado por seu objeto dinâmico apenas no sentido de que ele será assim interpretado. Ele depende, portanto, de uma convenção, um hábito ou uma disposição natural do seu interpretante ou campo do seu interpretante (aquilo de que o interpretante é uma determinação) (CP 8.335).
>
> [O símbolo] é um signo que se constitui como tal meramente ou principalmente devido ao fato de que ele é usado ou compreendido como tal, seja o hábito natural ou convencional, e independente dos motivos que originalmente governaram sua escolha (CP 2.307).
>
> O símbolo é um *representamen* cuja significância especial ou adequação para representar aquilo que ele representa não repousa em outra coisa senão no fato de haver um hábito, disposição ou outra regra geral efetiva de que ele seja assim interpretado (CP 4.447).

Embora, em algumas passagens, hábito e convenção sejam usados como sinônimos, há algumas diferenças sutis que devem ser discernidas. Para tal, é preciso levar em consideração que a noção peirceana de hábito, tanto quanto a de lei, é bastante original. Não é por acaso que ambos, lei e hábito, em alguns casos, podem ser tomados como sinônimos (ver Santaella 1999a,b). No papel que o hábito desempenha junto ao símbolo, vale a pena chamar a atenção para o fato de que hábitos podem ser inatos, incluindo, portanto, disposições naturais (CP 4.531). Se inclui a disposição natural, então nem todo símbolo é necessariamente convencional. Para confirmar essa constatação, há uma passagem bastante significativa de Peirce quando se auto-critica dizendo: 'notando que eu havia classificado sintomas naturais tanto entre os índices quanto entre os símbolos, restringi símbolos aos signos convencionais, o que foi um erro' (CP 2.340). Assim sendo, embora a imensa maioria dos símbolos seja, sem dúvida, convencional (CP 3.360), podem existir símbolos que dependem de hábitos naturais.

Neste ponto, compreender a originalidade da concepção peirceana de hábito pode contribuir para um melhor entendimento do próprio símbolo. Em uma certa medida, o hábito, de fato, é um conceito psicológico, no sentido em que se corporifica na mente humana. Mas não é apenas psicológico, pois 'hábitos são regras gerais às quais o organismo se submeteu' (CP 3.360). Além disso, organismos não precisam ser humanos. Há hábitos em organismos rudimentares, assim como há hábitos nas plantas e na própria natureza. Nessa medida, o conceito peirceano de hábito é muito geral e

abstrato. Trata-se de uma 'regra geral efetiva' (CP 4.447), isto é, de uma 'regra para a ação' (CP 5.397-98). Assim sendo, hábitos são ações que tendem a se repetir de acordo com padrões uniformes, sob condições específicas. Nesse nível de generalidade, o hábito é um sinônimo de lei adquirida ou natural. Quando ela é adquirida por um pacto coletivo, o hábito é convencional.

Aqui, recupera-se operacionalmente a noção de lei como 'regularidade no futuro indefinido' (CP 2.293) assim como as conseqüências que ela traz para o legi-signo simbólico, como se pode atestar nas passagens a seguir.

> O significado de um símbolo consiste no modo como ele pode nos levar a agir. É claro que esse 'como' não pode se referir à descrição de movimentos mecânicos que ele poderia causar, mas deve se referir à descrição da ação como tendo este ou aquele alvo (CP 5.135).
>
> A palavra não é uma coisa. Ela consiste na regra geral realmente operacional de que esses três traços (a palavra 'man') vista por uma pessoa que saiba inglês afetará sua conduta e pensamentos de acordo com uma regra. [...] O ser de um símbolo consiste no fato real de que algo será seguramente experienciado se certas condições forem satisfeitas. A saber, ele influenciará o pensamento e a conduta do intérprete (CP 4.447).

O hábito que o símbolo aciona na mente do intérprete implica em uma disposição para agir de um determinado modo, sob certas circunstâncias. Tal disposição encontra sua melhor expressão em uma proposição no modo condicional. Mas a questão ainda não se esgota aí. Se o signo simbólico é, em si mesmo, um legi-signo, essa lei é também uma regra geral ou hábito. Ou melhor, não apenas seu interpretante, mas o próprio legi-signo é também um hábito ou regra geral efetiva (CP 2.249). Só por isso ele é capaz de acionar, no campo do interpretante, uma regra interpretativa que, ao se corporificar na instância de um intérprete particular, produzirá uma associação de idéias gerais, uma regularidade associativa (CP 4.500), ou uma conexão habitual entre o signo e o objeto denotado (CP 1.369). No caso da linguagem verbal, vem daí o caráter geral, social da língua e, ao mesmo tempo, particular, individual do seu uso. As convenções lingüísticas só operam porque os indivíduos de uma comunidade inteira internalizaram hábitos de interpretação. São esses hábitos ou regras que Peirce chamou de interpretantes lógicos.

Como se pode ver, embora o fundamento da relação do símbolo com o seu objeto esteja, na maior parte das vezes, baseado em um caráter imputado (CP 1.558), ou seja, convencional, não é possível tratar a convencionalidade sem se considerar, de um lado, o legi-signo ou lei que determinará o interpretante (CP 2.292), de outro lado, sem se considerar o interpretante. É no interpretante que se realiza, por meio de regra associativa, uma associação de idéias na mente do intérprete (CP 2.299), associação esta que estabelece a conexão entre o signo e seu objeto. Daí Peirce repetir inúmeras vezes que o símbolo se constitui como tal apenas através do interpretante (NEM 4: 260).

Entretanto, nenhuma ocorrência interpretativa em um intérprete particular de um legi-signo simbólico pode esgotar a generalidade que lhe é própria. Vem daí a plasticidade do símbolo. Sua aptidão para a mudança. Tais mudanças são produzidas, quando ocorrem transformações no hábito interpretativo de um símbolo, pois as regras de interpretação, isto é, os interpretantes lógicos podem ser modificados. Por isso mesmo, o símbolo é um signo em crescimento nos interpretantes que ele gerará, no longo caminho do tempo (ver Short 1988).

Estando esclarecido o caráter geral, caráter de lei de toda a tríade, signo, objeto e interpretante do símbolo, há ainda uma questão crucial a ser discutida. Uma vez que as leis não têm existência concreta, de onde vem o poder denotativo do símbolo? Como podem as palavras se referirem àquilo que está fora delas?

Como já vimos, 'tudo que é geral tem seu ser nos casos que determina' (CP 2.249). O legi-signo depende de casos individuais para se materializar. O legi-signo simbólico toma corpo nesses casos individuais que, no ato mesmo de lhe dar corpo, conformam-se ao seu governo. Ele funciona, portanto, como uma lei ou regra para a formação de uma certa subclasse de sin-signos que são chamados de réplicas do legi-signo. A regra para a formação das réplicas envolve também a regra de interpretação dessas réplicas. Assim sendo, a réplica de um símbolo é um tipo especial de índice que age para aplicar a regra geral ou hábito de ação ou expectativa associada com o símbolo a algo particular (Short 1988). Para que essa aplicação a algo particular ocorra, é preciso haver casos existentes daquilo que o símbolo denota. Já foi discutido anteriormente que o objeto do símbolo é tão geral quanto ele próprio. Entretanto, há casos singulares aos quais ele se aplica. Como se aplica?

> Um símbolo em si mesmo é um mero sonho, ele não mostra sobre o que está falando. Precisa estar conectado a seu objeto. Para esse propósito um índice é indispensável. Nenhuma outra espécie de signo responderá a esse propósito. Que uma palavra estritamente falando não pode ser um índice é evidente a partir disto: uma palavra é geral, ela ocorre freqüentemente, e, todas as vezes em que ocorre, é a mesma palavra, e se ela tem algum significado como palavra, ela o terá todas as vezes em que ocorre; enquanto o índice é essencialmente um caso do aqui e agora, seu ofício sendo o de trazer o pensamento para uma experiência particular ou uma série de experiências conectadas por relações dinâmicas (CP 4.56).

É por isso que, no universo do discurso, há vários tipos de palavras, entre elas, as gerais, estritamente simbólicas, e as indiciais, como são os pronomes pessoais, demonstrativos, os advérbios de lugar etc. Estas últimas constituem o ingrediente indicial do símbolo, também chamadas de marcas enunciativas, cuja função é conectar o pensamento, o discurso, o signo geral a experiências particulares. Quando dizemos a palavra 'mulher', por exemplo, o referente ou objeto dessa palavra é um tipo geral que nenhum caso particular de mulher pode completamente recobrir. Mas, quando dizemos 'mulher brasileira', através do índice de lugar 'brasileira', indicação de nacionalidade, aí está o caso a que o geral se aplica. A incorporação de outros índices poderiam ir especificando cada vez mais o referente do discurso, como por exemplo, 'mulher brasileira dos anos 90' etc. Contudo, essa função conectora é tudo que o índice pode realizar, nela começa e nela acaba o papel que o ingrediente indicial do símbolo pode desempenhar. Por isso mesmo, falta ainda ser discutida uma outra interrogação bem mais crucial. De onde vem o poder do símbolo para significar? Conforme já demonstrei em outra ocasião (Santaella 1995: 172-175), a resposta para essa pergunta exige muita acuidade analítica.

Peirce distinguiu dois tipos de generalidade, de um lado, a generalidade objetiva ou referencial que está na capacidade de algo para representar uma pluralidade de objetos. De outro lado, a generalidade subjetiva, que Ransdell (1966: 158-160) chama de generalidade entitativa para indicar que ela é qualificadora. Qualquer coisa é entitativamente geral, se o seu modo de ser não é o de um individual (CP 5.429, 1.420). Essa generalidade entitativa, daquilo que não é um individual, foi então dividida por Ransdell em qualitativa e nômica. A primeira 'é de uma espécie negativa e per-

tence ao que é potencial como tal, e isso é peculiar à categoria da qualidade'. A segunda é daquela espécie positiva que pertence à necessidade condicional e esta é peculiar à categoria da lei (CP 1.427). Ransdell diz (1966):

> Não conheço outro modo de caracterizar esses dois tipos de generalidade entitativa, a não ser notando que elas correspondem à primeiridade e terceiridade peircianas, o que pode ser ilustrado do seguinte modo. De um lado, não faz sentido perguntar 'Onde e quando é a vermelhidão?', e vermelhidão (a forma, primeiridade, qualidade) é geral precisamente por essa razão. Por outro lado, faz bastante sentido perguntar onde e quando algo é vermelho; mas para essa questão duas respostas são possíveis. Pode-se dizer 'Isto, aqui e agora, é vermelho', e isso que está sendo denotado seria um individual, e portanto, não geral. Ou pode-se dizer: 'Algo (isto é, qualquer coisa) será vermelho, quando tais e tais condições forem preenchidas', e esta resposta não faria referência a qualquer coisa individual, mas denotaria uma *regularidade* ou *classe* de casos dos quais seria verdadeiro dizer de qualquer um, que seja dado, que 'Este, aqui e agora, é vermelho', sendo essa classe definida por condições específicas. Nesse caso, o que é denotado seria nomicamente geral.

Há, portanto, dois modos de generalidade: (1) objetiva ou referencial; e (2) subjetiva ou entitativa, esta subdividida em (2.1) qualitativa e (2.2) nômica. O sin-signo indicial é o único tipo de signo que está desprovido de generalidade. Ele sempre indica, aponta para individuais ou coleção de individuais. O ícone apresenta uma generalidade entitativa do tipo qualitativo. O símbolo, por sua vez, possui tanto a generalidade referencial, geral, quanto a entitativa de tipo nômico, isto é, a generalidade que pertence à necessidade condicional. Mas, uma vez que o símbolo contém dentro de si elementos de iconicidade e elementos de indicialidade, o símbolo funciona como síntese de todas essas dimensões. Como isso se processa é o que será visto a seguir.

Ao retomar as noções lógicas tradicionais de compreensão (profundidade) e extensão (aplicação), Peirce considerou-as como as duas propriedades semióticas do símbolo. O nome que deu a elas foi variado, tais como significação, conotação para a profundidade e denotação para a extensão. Enquanto a denotação, extensão ou aplicação, isto é, o poder aplicativo, referencial do símbolo corresponde ao seu ingrediente indicial, a significação, conotação ou profundidade corresponde ao seu ingrediente icônico. Qual se-

ria, então, o ingrediente propriamente simbólico do símbolo? Essa pergunta é procedente porque, se o símbolo se caracteriza como o signo mais genuinamente triádico, é de se esperar que seus ingredientes sejam três. De acordo com Ransdell (1966: 183), Peirce não nomeou explicitamente três propriedades provavelmente porque o ingrediente icônico e o simbólico estão tão profundamente atados que a distinção do papel desempenhado por cada um deles exige penetração analítica.

Já foi discutido que, para ligar 'o pensamento a uma experiência particular ou uma série de experiências conectadas por relações dinâmicas' (CP 4.56), o símbolo precisa de índices. Assim, o poder de referência, poder indicativo do símbolo vem de seu ingrediente indicial. Entretanto, o índice está desprovido do poder de significar. Por isso mesmo, para significar, o símbolo precisa de um ícone. Nesse caso, não se trata de um ícone tout court, mas de um tipo especial de ícone, a saber, um ícone que está atado a um ingrediente simbólico. Esse ingrediente, ou parte-símbolo, Peirce chamou de conceito; a parte-ícone, ele chamou de idéia geral. Para Ransdell (ibid.: 184), o conceito é o sentido e a idéia geral é a significação. A parte-símbolo, conceito ou sentido, corresponde ao hábito geral e não atualizado. A parte-ícone ou idéia geral é aquilo que atualiza o hábito produzindo a significação. É por isso que Peirce repetiu tantas vezes que o símbolo significa por meio de um hábito e de uma associação de idéias. Hábito não é tomado no sentido psicológico-prático, mas em um sentido similar àquele que Kant deu para o termo esquema ou regra, quando discutiu os esquematas que estão subjacentes aos nossos conceitos sensíveis puros, muito diferentes das imagens dos objetos (Ransdell ibid.: 167-171). A distinção peirceana entre o conceito ou hábito e a idéia geral está bem clara na citação a seguir:

> Uma idéia, que pode grosseiramente ser comparada a uma fotografia composta, ganha vividez, e essa idéia composta pode ser chamada de idéia geral. Não é propriamente um conceito; porque o conceito não é, de modo algum, uma idéia mas um hábito. Porém, a ocorrência repetida de uma idéia geral e a experiência de sua utilidade, resulta na formação de um hábito ou fortalecimento daquele hábito que é o conceito; ou se o conceito já é um hábito cuidadosamente compacto, a idéia geral é a marca do hábito (CP 7.498).

Assim, nossa idéia geral, digamos, de um gato, por exemplo, seria a fusão resultante de imagens decorrentes das situações re-

petidas de experiências sensórias mais determinadas e muito diferenciadas de gatos particulares. A idéia geral seria a *gestalt*, forma ou unidade imediatamente percebida, isto é, ícone, um geral entitativo de tipo qualitativo. A parte-ícone do símbolo é, portanto, a atualização do conceito, a concreção do conceito ou hábito que é, por sua vez, um geral objetivo ou referencial tanto quanto subjetivo ou entitativo do tipo nômico. Esse é o ingrediente autenticamente simbólico do símbolo, tão geral que, sem o auxílio de índices, para particularizar sua referencialidade, e do ícone, para concretizar sua generalidade nômica, ele, o símbolo, seria totalmente impotente para informar e significar qualquer coisa. Vejamos, assim, como o conceito e a idéia geral funcionam quando o símbolo é uma palavra.

Na passagem 2.292, citada acima, Peirce afirmou que a palavra como legi-signo, tipo geral, incorpora-se em existentes. Esses existentes, por sua vez, devem se conformar ao ser real da palavra. Segundo Ransdell (ibid.: 185), esses existentes não são réplicas em si, mas sim ocorrências individuais da interpretação das réplicas. Quer dizer, 'o existente em questão é a atualização do conceito pelas réplicas, atualização esta que toma a forma da manifestação de uma idéia geral'. Isso não significa que, ao ouvir, por exemplo, a palavra 'mulher' , a imagem de uma mulher salta em nossa cabeça. Segundo Ransdell, o que vem à mente é 'um conjunto antecipatório' ou *gestalt* resultante de uma mistura de dados perceptivos reais e imaginários.

Como todos os símbolos, a palavra também contém o ingrediente propriamente simbólico do símbolo, a saber, o conceito ou hábito. Entretanto, Peirce afirmou que a palavra e o conceito são regras gerais. Há aí duas regras, portanto. Para essa dualidade, Ransdell (ibid.: 187) fornece uma explicação muito clara. A regra, que é a palavra, é puramente intralingüística, ou seja, regra que determina as combinatórias permitidas e proibidas para a palavra no sistema da língua. Já a regra ou lei que é o sentido ou hábito é a regularidade do conceito. As línguas humanas relacionam, por meio de associações de idéias, as regras intralingüísticas com as regras do conceito. É por isso que o símbolo 'homem' ou o símbolo 'seis' não são as palavras 'homem' e 'seis' , mas sim o conceito de homem e seis nas suas manifestações de fato com as palavras 'homem', 'homme', 'hombre', 'man' etc. e com as palavras 'seis', 'six', 'sechs', 'zes' etc. As réplicas das palavras atualizam o conceito tanto na sua

manifestação denotativa, aplicativa (índice) quanto na sua manifestação icônica. Como atualização do conceito que constitui o sentido do símbolo, o ícone é uma idéia geral que o símbolo produz ao se concretizar em uma réplica. Senão vejamos:

> Um homem, caminhando pelo passeio junto com uma criança, levanta o braço, aponta e diz: 'Ali vai um balão'. Apontar é parte essencial do símbolo, sem o que este não veicularia informação. A criança, entretanto, pergunta: 'O que é um balão?', e o homem responde: 'É algo como uma grande bolha de sabão', tornando a imagem parte do símbolo. Assim, embora o objeto integral de um símbolo, isto é, seu significado, tenha uma natureza de lei, ele deve denotar um individual e expressar um caráter (CP 2.293).

Estando os ingredientes indiciais e icônicos do símbolo explicitados, falta ainda discutir os tipos de interpretantes que o legi-signo simbólico está apto a produzir. Todos os exemplos de legi-signos simbólicos dados até agora foram palavras isoladas. Nesse caso, o interpretante tende a representar esse signo rematicamente, o que o enquadra, portanto, na classe dos legi-signos simbólicos remáticos. O exemplo fornecido por Peirce desse tipo de signo sintetiza com perfeição os ingredientes do símbolo, isto é, o conceito ou hábito que corresponde ao ingrediente propriamente simbólico, a idéia geral ou ingrediente icônico e a aplicabilidade ou ingrediente indicial. Embora longa, essa citação (CP 2.261) merece ser transcrita sem cortes.

> Um símbolo remático ou rema simbólico (exemplo, um substantivo comum) é um signo relacionado com seu objeto por uma associação de idéias gerais, de maneira tal que sua réplica desperta uma imagem no espírito, imagem que, devido a certos hábitos ou disposições daquele espírito, tende a produzir um conceito geral, sendo a réplica interpretada signo de um objeto que é um caso daquele conceito. Assim, o símbolo remático ou é ou muito se assemelha ao que os lógicos chamam de termo geral. O símbolo remático, como qualquer Símbolo, participa necessariamente da natureza de um tipo geral e é, assim, um legi-signo. Sua réplica, todavia, é um sin-signo indicativo, remático de tipo especial, no sentido de que a imagem que sugere ao espírito atua sobre um símbolo já naquele espírito, para dar surgimento a um conceito geral. Nesse sentido, difere de outros sin-signos indicativos, remáticos, inclusive daqueles que são réplicas de legi-signos indicativos, remáticos. Assim, o pronome demonstrativo 'aquele' é um legi-signo, por ser de um tipo geral; mas não é um Símbolo, pois ele não significa um conceito geral. Sua réplica dirige a atenção para um objeto singular e

> é um sin-signo indicativo remático. Uma réplica da palavra 'camelo' é também um sin-signo indicativo, remático por ser realmente afetada, como conseqüência do conhecimento de camelos, comum a quem fala e a quem ouve, pelo camelo real que denota, ainda que este não seja individualmente conhecido de quem ouve, e é por essa conexão real que a palavra 'camelo' desperta a idéia de um camelo. O mesmo é verdadeiro em relação à palavra 'fênix'. Embora a fênix não exista realmente, reais descrições da fênix são bem conhecidas de quem fala e de quem ouve e, assim, a palavra é realmente afetada pelo objeto denotado. As réplicas dos símbolos rem áticos são muito diferentes não apenas dos sin-signos indicativos rem áticos ordinários, mas também diferem destes as réplicas dos legi-signos indicativos rem áticos. Com efeito, a coisa denotada por 'aquele' não afeta a réplica da maneira mais simples e direta como, por exemplo, o tilintar da campainha do telefone é afetado pela pessoa que, no outro extremo da linha, deseja estabelecer comunicação. O interpretante de um símbolo remático com freqüência o representa como legi-signo icônico, e com efeito, e em reduzida proporção, ele participa da natureza de ambos.

Das palavras isoladas, passamos para as proposições. Neste caso, o interpretante tende a representar o signo como um dicente, o que o enquadra na classe de legi-signo simbólico dicente. Um símbolo dicente, ou proposição ordinária, é um signo que se relaciona com seu objeto por uma associação de idéias gerais e que age como um símbolo remático, exceto pelo fato de que seu pretendido interpretante representa o símbolo dicente como sendo, com respeito àquilo que ele significa, realmente afetado por seu objeto, de sorte que a existência ou lei que ele faz surgir no espírito deve estar efetivamente relacionada com o objeto indicado. Assim, o interpretante contempla o símbolo dicente como um legi-signo indicativo, dicente; e se isso for verdadeiro, partilha dessa natureza, embora aí não se esgote. À semelhança do sin-signo dicente, ele é composto, de vez que necessariamente envolve um símbolo remático (e assim é para seu interpretante um legi-signo icônico) para expressar-lhe a informação, e um legi-signo indicativo remático para assinalar a matéria daquela informação. Contudo, a sintaxe desses é significativa. A réplica do símbolo dicente é um sin-signo dicente de tipo especial. Facilmente percebemos ser isso verdadeiro quando a informação que o símbolo dicente veicula é relativa a um fato concreto. Quando aquela informação diz respeito a uma lei real, ele não é verdadeiro na mesma extensão. Com efeito, um sin-signo dicente não pode veicular informação de lei.

É, conseqüentemente, verdadeiro em função da réplica de tal símbolo dicente apenas na medida em que a lei tem seu ser traduzido em exemplos.

Das proposições, passamos para o nível do discurso. Neste caso, o interpretante tende a representar o signo como um argumento, o que o enquadra na classe de legi-signo simbólico, argumental, a mais abstrata entre todas as classes de signos. Conforme já explicitei em outra ocasião (Santaella 1995: 192), o argumento é um signo que é interpretado como um signo de lei, regra reguladora ou princípio guia, ou melhor, 'é um signo cujo interpretante lhe representa o objeto como sendo um signo ulterior, por meio de uma lei', a saber, a lei segundo a qual 'a passagem de todo o conjunto das premissas para as conclusões tende a ser verdadeira' (CP 2.203). Há mecanismos que derivam conclusões válidas de premissas, mas Peirce não chamou esses processos de argumentos. Um argumento deve ser compreendido por seu interpretante como derivando validamente uma conclusão de suas premissas porque ele pertence a uma classe de inferências possíveis que se conformam com um princípio guia. Esse é o princípio de funcionamento do silogismo.

O argumento deve ter um caráter geral, o que significa que só legi-signos simbólicos podem ser argumentos. As réplicas dos argumentos são sin-signos dicentes. Peirce dividiu os argumentos em abdutivos, indutivos e dedutivos. A partir dessa divisão pode-se concluir que o silogismo é apenas a manifestação mais formal do argumento. Contudo, há outros tipos de manifestação não tão radicalmente dedutivas de modo que se pode pensar na possibilidade não apenas de discursos dedutivos, mas também indutivos e mesmo abdutivos. Neste ponto, a pergunta proposta no título deste capítulo, o que é o símbolo, parece ter atravessado todas as esferas com que a teoria de Peirce nos permite respondê-la.

Referências

PEIRCE, CHARLES S. (1994 [1866-1913]). *The Collected Papers of Charles S. Peirce*. Electronic edition reproducing Vols. I-VI [Ed. Hartshorne, C. & Weiss, P., Cambridge: Harvard University, 1931-1935], Vols. VII-VIII [Ed. Burks, A. W., Cambridge: Harvard University, 1958]. Charlottesville, Intelex Corporation.

__. 1976. *New Elements of Mathematics by Charles S. Peirce*. Eisele, C. (ed.). The Hague: Mouton.

RANSDELL, JOSEPH. 1966. *Charles Peirce:* The Idea of Representation. Tese de doutoramento. Columbia University.

__. 1983. *Peircean Semiotics.* [Manuscrito inédito].

SANTAELLA, LÚCIA 1995. *Teoria geral dos signos. Semiose e autogeração.* Ática, 2a. ed. (2000). Ed. Pioneira.

__. 1999a. Caos, acaso e lei em Peirce: uma nova causalidade. Em *Caos e Ordem na Filosofia e Ciências*, L. Santaella & J.A. Vieira (orgs.). Face 2: 26-37.

__. 1999b. A new causality for the understanding of the living. *Semiotica* 127 (1-4): 497-519.

SANTAELLA, LÚCIA e NÖTH, WINFRIED. 1998. *Imagem. Cognição, Semiótica, Mídia.* Iluminuras, 3a. ed. 2002.

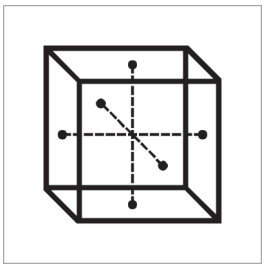
SAVAN, DAVID. 1976. *An Introduction to C. S. Peirce´s Full System of Semiotic.* Victoria College of the Unversity of Toronto.

SHORT, THOMAS. 1988. The growth of signs. *Cruzeiro Semiótico* 8: 81-87

COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA



CAPÍTULO 5

ASPECTOS METODOLÓGICOS DA SEMIÓTICA COMPUTACIONAL

Alexander Mehler

INTRODUÇÃO

A Semiótica Computacional (SC) surge onde se interseccionam semiose e computação (Clarke 2001). Qualquer definição da SC tem como ponto de partida a noção de signo e de processo sígnico. Neste capítulo, não tentaremos reconstruir este fundamento. Vamos esboçar algumas de suas implicações metodológicas. Não obstante, vamos esquematizar o que parecem ser estes fundamentos. Considerando que os objetos de interesse do autor são sistemas de discurso de linguagem natural, este capítulo é necessariamente uma triagem das linhas de pensamento que acreditamos que precisam ser levadas em consideração para esta tarefa:

> I. Peirce propõe uma semântica dinâmica, relacional, que descreve 'significado' como resultado de um processo contínuo de interpretação de signos resultando numa constituição/ modificação de disposições comportamentais, e restringindo o uso

de signos exatamente nestes processos. Sem seguir a noção triádica de signo de Peirce, é este tipo de 'circularidade' que acreditamos ser crucial para uma fundamentação semiótica da SC, isto é, o fato de que signos não somente participam do processo sígnico na base de disposições (regularidades de uso), mas também podem, como resultado da sua participação, mudar estas disposições. As implicações desta noção para o conceito lingüístico de *regra*, ou, mais geral, de *regularidade*, são múltiplos, desde que a perspectiva dinâmica de Peirce não permite concebê-los como entidades estáticas.[1]

II. Enquanto Peirce não faz distinção entre as regularidades sintagmática e paradigmática, especialmente em relação a unidades textuais no nível de estruturas argumentativas, um tópico central de sua filosofia de signo, esta distinção, que confronta a distinção entre *sistema de texto* e *sistema de linguagem*, é fundamental para a glossemática de Hjelmslev (1969). Conseqüentemente, Hjelmslev pode descrever a lingüística como uma abordagem formal, dedutiva, que, começando de um texto não analisado como um todo, tenta reconstruir o sistema de escolhas que define o sistema de linguagem, e as realizações destas escolhas que definem as instâncias textuais. Desconsiderando a abordagem dedutiva da glossemática, a dicotomia da sintagmática e da paradigmática é vista como essencial para o fundamento semiótico da SC.

III. Em contraste com a noção estática de Hjelmslev do sistema de linguagem, é a dinâmica 'texto' — 'constituição do sistema de linguagem' que é acentuada na Lingüística Funcional Sistêmica, de Halliday (LFS) (1977), com ênfase na sensibilidade ao contexto dos processos lingüísticos. Neste sentido, um texto sendo produzido/recebido como uma unidade de discurso, por pelo menos um participante de uma comunidade, sempre tem ao menos dois (tipos de) contextos: o sistema de escolhas lingüística e semântica que estão sobre ele, e o (tipo de) contexto social que LFS diferencia em relação à variedade, de acordo com fatores situacionais como campo, teor e modo (cujas recorrentes combinações são descritas como *registros*) e de acordo com o estágio de interações sociais (descrita como *gêneros*) (Halliday 1977, Martin 1992).[2] Como conseqüência desta sensibilidade ao contexto, a lingüística não pode ser conduzida como uma disciplina puramente

> dedutiva, mas necessariamente constrói uma análise qualitativa, e também quantitativa, de aspectos sincrônicos e diacrônicos da dinâmica das estruturas lingüísticas. Não obstante, a sintagmática e a paradigmática continuam termos constitutivos em LFS, mas são agora — conforme Peirce — analisados de uma perspectiva dinâmica. Um co-envolvimento dinâmico de texto e sistema, bem como seu aspecto contexto-sensitivo, especialmente seu entrelaçamento em *sistemas sociais*, é visto como essencial para o fundamento semiótico da SC.

A unificação destes objetivos aponta para uma abordagem estrutural e também procedimental: é estrutural no sentido que modela signos em relação à sua incerteza sintagmática e regularidades paradigmáticas; é procedimental no sentido que não só modela estas regularidades como resultado, mas também como uma condição prévia de processos sígnicos. Como entidades dinâmicas, estes processos são necessariamente modelados por meio de procedimentos.

A síntese de aspectos estruturais e dinâmicos previne certas dificuldades: primeiro, a suposição sobre a existência de unidades de significado atômicas pode ser abandonada. Em vez de proclamar um nível adicional, empiricamente não-observável de átomos semânticos, esta abordagem recorre a uma análise do potencial variável de signos para interpretação. Segundo, o sistema de linguagem não é analisado separado de sistemas de discurso, nem de contextos sociais: qualquer regularidade explorada está sempre associada a unidades de co-variação contextual, que são exploradas como entidades semióticas, elas mesmas.

São focalizadas algumas implicações metodológicas da Semiótica Computacional (SC) a seguir.

Simulação, realização e emulação

A abordagem semiótica para a 'cognição situada' (Rieger 2001, Strohner 1995, Varela 1993) implica que sistemas de processamento de informação têm sua *endo-visão* constituída por meio de processos sígnicos. Tais signos são, necessariamente, parte de um sistema de linguagem que é constituído por um sistema social. A modelagem de tais sistemas semióticos exclui, necessariamente, qual-

quer identificação da *exo-visão* do projetista do modelo com a *endo-visão* do sistema. A identificação problemática da endo-visão e exo-visão é característica para a abordagem cognitiva e associativa (Varela 1993), mas não o é para a abordagem semiótica (Rieger 2001). Ao contrário, a abordagem semiótica objetiva modelar esses processos como resultado de sistemas semióticos, constituindo seus próprios ambientes. Para lançar alguma luz ao termo 'Semiótica Computacional', nos referimos à distinção de Pattee (1988) entre *simulação* e *realização*, como uma base para distinguir diferentes abordagens para a SC. Isto é feito em complementação ao termo emulação.[3]

> I. Seguindo termos da teoria clássica de modelo, *simulações* são necessariamente homomórficas; descrições formais do seu original — eles *simbolicamente* estão para as entidades que modelam. Como um *differentium specificum*, baseiam sua função na reconstrução e predição do seu original, respectivamente. Simulações são dinâmicas, modelos procedimentais de funções selecionadas de seus originais. Assim, elas necessariamente incluem tempo como um parâmetro constitutivo. Enquanto, homomorficamente, as simulações mapeiam qualquer noção de tempo (físico, biológico ou semiótico), elas não *realizam* estas funções. As funções simuladas, e estruturas nas quais operam, permanecem ontologicamente separadas das suas contrapartes formais. Porém, como modelos procedimentais, as simulações introduzem um nível adicional de falsificação: são falsificáveis com relação à sua organização procedimental, e com relação às reconstruções e predições que elas produzem.[4]
>
> II. *Realizações* são modelos materiais que *implementam* certas funções de seus originais, e que existem na mesma área ôntica destes originais. Como simulações, realizações têm uma organização procedimental, mas seus exemplos servem para realizar as funções em questão. Uma pergunta central colocada nos estudos de Vida Artificial é se esta realização é possível apenas no mundo 'real' (físico), ou também em mundos artificiais. Em todo caso, realizações são avaliadas com base na função que realizam, não com relação ao benefício das medidas (correspondências) que executam. Se, por exemplo, 'voar como as libélulas' é a função a ser realizada, um

helicóptero pode ser uma boa realização desta função, mas é um modelo ruim de libélulas.

III. Acrescentamos um terceiro termo: *emulação*. Neste capítulo, emulações são, sob a perspectiva de 'processo', mapeamentos entre dois modelos de simulação ou realização. Emulações buscam transformar o sistema que está sendo emulado em outro sistema, mas igualmente efetivo, em termos funcionais, ou um bom meio, em termos de simulação. Se, por exemplo, um sistema operacional é emulado em outro sistema operacional, é irrelevante se o sistema emulador realmente executa as funções do sistema emulado, do modo como o segundo o faz. É justamente o efeito igual que importa quando as funções correspondentes são utilizadas no meio emulado.

Em termos formais podemos distinguir assim simulações, realizações e emulações: seja G um conjunto de sistemas sendo modelados, F um conjunto de funções de sistema ('voar', 'falar') de sistemas fora de G, e M um conjunto de sistemas artificiais, ou modelos utilizados para modelar/medir sistemas em G. Usando estes termos, uma simulação S pode ser descrita como uma associação de um sistema e um modelo, isto é, $S \in G \times M$, em que o sistema é mapeado em um modelo por meio de um homomorfismo. Por outro lado, uma realização $R(F_1)$ é uma associação de um subconjunto de funções $F_1 \subseteq F$, um sistema artificial $s \in G \cap M$, servindo as funções em F_1, formalmente, $F_1(s)$, e um subconjunto de sistemas servindo as mesmas funções, isto é, $R(F_1) \in \{ s \in G \cap M \mid F_1(s) \} \times \{ r \in G \setminus M \mid F_1(r) \}$. Finalmente, uma emulação E é uma associação entre elementos em M, isto é $E \in M \times M$, onde, no caso de simulações, um homomorfismo existe entre ambos sistemas associados pela emulação.[5]

Se concebermos a SC como uma disciplina que constrói modelos procedimentais de processos baseados em signos, a decisão de quando estes modelos são simulações, realizações, ou algo entre um e outro, é mais delicada do que se pode esperar. Isto pode ser esboçado da seguinte forma: em um sentido estrito, existe uma interpretação 'negativa' de simulações como modelos de interações entre os sistemas e seus ambientes, em que o designer do modelo pressupõe os resultados das medições das operações que constituem os ambientes dos sistemas. 'Pressupor' significa que o designer enumera categoricamente o universo de todos os possíveis contex-

tos do sistema em referência a uma semântica composicional que opera nas unidades atômicas de representação de contexto (relações, localizações, variáveis de tempo, etc), reduzindo, assim, a relação de significado do sistema a uma *função de significado* clássica – por exemplo, para o sistema *s*, o significado do signo α no contexto X_{i1} é m_{j1}, o significado do signo β no contexto X_{i2} é m_{j2}, etc.). Rieger (2001: 167) caracteriza esta visão como segue:

> Para tentar compreender (condições de possibilidade) a compreensão de signos e significados não se pode cair no processo *simulativo* de estruturas (símbolo) cujo estado representacional é declarado por meio de uma semântica pré-estabelecida (conhecida pelo modelador, acessível para o modelo, mas não obrigatória para o sistema modelado). Ao invés disso, a modelagem de processos de constituição de significado terá de realizar as funções implementadas e operacionais em um sistema de processamento de informação capaz de produzir alguma estrutura, em um processo de auto-organização, representacional de alguma outra coisa, e que também permita identificar para o que esta estrutura está para.

Em outras palavras, realizações não precisam e, por conta da complexidade dos ambientes em que estes *sistemas implementacionais* operam, nem mesmo permitem a enumeração simbólica de seus possíveis contextos. Seu universo contextual não pode ser extensamente enumerado, nem, como acreditamos, intencionalmente especificado em termos estáticos, mas apenas definido procedimentalmente. Em relação à abordagem semiótica, isto significa que qualquer representação estática de significado, puramente simbólica, como uma função entre conjuntos enumeráveis, o que pressupõe conhecimento sobre todos os possíveis referentes de todos os signos modelados, em contextos possíveis igualmente pressupostos, tem que ser substituída por uma 'instrumentação procedimental' da simulação/realização. Isto permite que, autonomamente, se descubra o que é um contexto relevante e que referentes, de quais signos produzidos/recebidos pelo sistema, são adequados em que contextos.

Mas como qualificar um modelo em que ambos, o sistema e seu ambiente, são implementados em um computador? Como um sistema pode ser avaliado se tanto ele quanto seu ambiente são modelos procedimentais, necessariamente operando em uma 'física' distinta de seu original. Em que sentido tal sistema pode ser chamado uma realização? Ele é só uma simulação? Se ele é, as funções

que executa não são realizadas e, conseqüentemente, o sistema *não pode* ser visto como *emulando* um sistema real que executa as funções. Qualquer abordagem relacionada a modelos de realizações sistema-ambiente implica que estes modelos, como um todo, executam (emulam) as mesmas funções (mas não necessariamente da mesma maneira) de seus originais. Claramente, no caso de funções sensório-motoras, isto é impossível. Mas o que dizer sobre funções cognitivas? Elas são distinguíveis de suas realizações corpóreas, funcionalmente, e equivalentemente, implementáveis em um computador? É bem sabido que computadores são ferramentas universais para simulação de processos que podem ser formalmente representados como funções recursivas. Este lugar comum significa que qualquer modelo computacional se restringe aos limites destas funções simbolicamente representáveis? Se é assim, seria impossível uma realização computacional de funções incorporadas, porque estes modelos não abandonam a esfera simbólica.

A dificuldade para achar respostas adequadas a estas perguntas indica que o espectro coberto pelos modelos discutidos inclui um tipo de modelo de simulação computacional que, como simulação, *não* realiza qualquer função cognitiva mas simula seus resultados, *bem como* sua organização procedimental *sem* confiar em qualquer semântica pré-estabelecida e em qualquer modelo de contexto. Mas, ao compará-las às realizações, estes modelos autonomamente executam operações de medidas, nos seus ambientes artificiais, ao fim das quais eles semioticamente produzem representações contextuais, de acordo com suas próprias estruturas, necessidades, etc. Entretanto, este sistema nunca se torna inteligente ao simular funções cognitivas, ele não realiza as funções que simula.[6] Assim, é duvidoso falar de emulações, no caso destes modelos, já que as funções que eles supostamente emulam não são realizadas por seus equivalentes modelos teóricos. Tome, uma vez mais, o exemplo do vôo: um simulador de vôo não emula o vôo, já que não há ninguém voando, ao usar o simulador. Além disso, se tudo, avião, ambiente, e piloto, são parte de um modelo computacional, não há razão para postular que o modelo emularia o vôo, de um modo funcionalmente equivalente.

Vamos revisar algumas abordagens, na área de SC, e campos relacionados, de acordo com os termos apresentados. Mais concretamente, vamos perguntar que tipo de modelagem elas realizam (simulações, emulações ou realizações de sistemas semióticos):

I. *Semiótica dos Computadores:* referindo-se ao estruturalismo de Hjelmslev, Andersen (1990) descreve os princípios semioticamente fundados para a análise de sistemas computacionais, entendidos como sistemas semióticos. Ele procura analisar os aspectos semióticos de sistemas de softwares, bem como de processos de desenvolvimento de sistemas e projetos de interface homem-computador. Seguindo o título de seu livro inaugural, e separando-o das abordagens descritas abaixo, chamamos a abordagem de Andersen de *Semiótica dos Computadores*. Ela trata de computadores, e de sistemas relacionados, mas não tem implicação relativamente às questões sobre simulação e realização.

II. Semiose Artificial é comparável à abordagem em *Vida Artificial* (Pattee 1988) e *IA forte* (Searle 1980): proposta de *realizar*, ou *emular*, comportamento inteligente por meio de sistemas computacionais artificiais. Como sistemas animados, eles experimentam, por exemplo, seu ambiente e autonomamente/ inteligentemente interagem com ele (Döben-Henisch 2002), e até aprendem linguagens. Para acentuar a relação desta abordagem com a IA forte propomos chamá-la de *Semiose Artificial*.

III. *Abordagem SECSE:* além da *Semiótica de Computadores e Semiose Artificial*, um campo adicional de SC pode ser destacado. Ao invés de tentar realizar sistemas semióticos artificiais, ele segue a abordagem de simulação esboçada acima. A abordagem de Andersen (1990), que descreve 'gêneros' como sistemas auto-organizados, pode ser vista como um exemplo. Comparável a sistemas computacionais de previsão do tempo, que não realizam sistemas climáticos (nem produzem temperatura, chuva, ou vento), mas simulam mudanças climáticas sem pré-estabelecer todos os estados, e mudanças de estados, a implementação de estudos nesta área não abrange sistemas semióticos, mas simula sua organização procedimental. Como conseqüência, a interpretação final, e avaliação das estruturas produzidas, por estes sistemas, permanecem do lado do construtor/usuário do modelo. Ao contrário da *semiose artificial*, onde o sistema artificial pode operar como um agente, juntamente com outros sistemas cognitivos animados, a interpretação não pode ser delegada ao suposto sistema autônomo. Para dar a esta abordagem um

nome, a chamaremos de *Semiótica Computacional em um Sentido Estrito (SECSE)*.

É esta última abordagem que nos interessa aqui. Por conta desta escolha, utilizaremos os termos 'Semiótica Computacional' (SC) e 'SECSE' intercambiavelmente.

MODELOS PROCEDIMENTAIS

Conforme uma terminologia de Marr (1982), a SC se concentra em modelos *procedimentais* de sistemas de signo. Ela questiona a modelagem de processos que produzem as estruturas semióticas em consideração. Esta mudança do foco de interesse está alinhada a uma preferência metodológica por modelos computacionais como meios para *simular* processos sígnicos: enquanto a análise semiótica formal (por exemplo, lingüística algébrica) está primariamente baseada em cálculo, para uma descrição de aspectos estruturais de signos, a SC integra ambos — a análise formal de aspectos estruturais e dinâmicos. Isto é apresentado na figura (1), onde as fases do modelo conceitual, cálculo e descrição, ordenadas de forma decrescente de abstração, são alinhadas por procedimentos, algoritmos (realizações de procedimentos) e implementações (realizações de algoritmos). Enquanto o primeiro ramo mantém o foco primariamente em *descrições* de acordo com os critérios de completude, consistência, e simplicidade, como declarado em lógica, é a reconstrução das estruturas semióticas sob observação que está sob no foco da SC. Esta *reconstrução* está associada à especificação de algoritmos, possivelmente aumentando as classes de algoritmos conhecidos, e suas implementações como programas de computação.

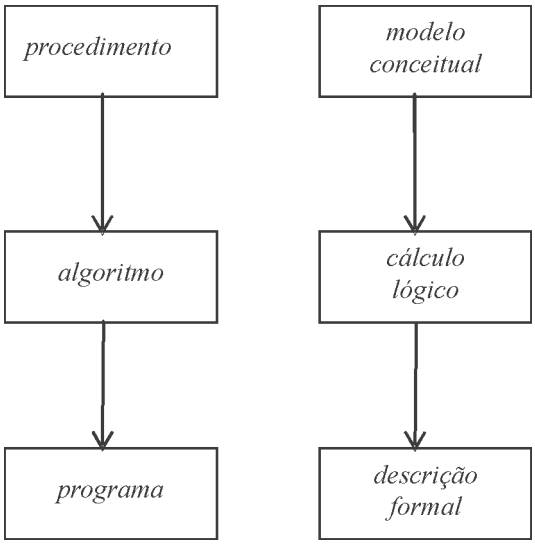


Figura 1: Níveis de abstração procedimental e conceitual.

Uma questão central levantada na SC é a dinâmica inter-relação de signos e contextos situacionais, cognitivos e sociais de sua produção-recepção: não apenas que o mesmo signo possa ser, dependendo da variação de contextos, diferentemente interpretado (polissemia), mas que seu uso contínuo tem potencial para modificar as condições de sua atualização (aprendizado). Apesar da proximidade com a área de aprendizagem de máquinas, a SC parte de abordagens que modelam processos de aprendizagem *convergente*, com recurso para um conhecimento anterior. Modelos de aprendizagem de SC são necessariamente não-supervisionados, se referem a conhecimento procedimental de como adquirir conhecimento, e permitem evoluir o objetivo da aprendizagem, além de fazer o processo de aprendizagem *divergir* temporariamente.[7] Esta saída é refletida pela restrição da *interpretabilidade procedimental*: a SC exige que a organização procedimental de modelos computacionais semióticos sejam interpretáveis no sentido de possuírem originais semióticos.

Apesar deste momento de 'naturalização' metodológica, a SC *não* tenta *realizar* máquinas semióticas (comparável a qualquer tipo de IA rígida), mas construir sistemas de informação baseados em modelos de signo que reflitam a dinâmica de processamento semiótico de informação para melhor se ajustar às necessidades dos seus usuários humanos.

Vamos sumarizar a concepção de SC aqui esboçada: um modelo computacional semiótico é um modelo simulativo, procedimental, que inclui formatos representacionais (i) para a representação do sistema, contexto e processo, formatos que (ii) permitem modelar as dinâmicas do sistema e mudanças de contexto, bem como de vários aspectos de incerteza informacional, (iii) não é apenas algoritmizável, mas também implementável, e (iv) tem a organização procedimental que é semioticamente interpretável — não a reprodução de um certo efeito pela realização/emulação de uma certa função, mas o modo como o sistema original, que está executando sua função, é simulativamente modelado. (v) Modelos computacionais semióticos servem para implementar um tipo de mecanismo de aprendizagem de máquina onde o sistema de aprendizagem evolui sua capacidade bem como seus objetivos. Assim, modelos computacionais semióticos necessariamente realizam um tipo de processamento de informação *explorativa*, onde sistema e ambiente de aprendizagem evoluem, mutuamente, não sendo pré-estabelecidos pelo designer. Modelos computacionais semióticos são formalizados em termos algorítmicos. As estruturas produzidas por suas implementações estão associadas à informação contextual que sustentam seus processamentos. Modelos computacionais semióticos são falsificáveis relativamente aos seus escopos teóricos, suas exigências procedimentais, e as estruturas que eles produzem (reconstroem/predizem).

Notas

[1] Neste sentido, a filosofia do signo de Peirce pode ser vista como uma possível base de corpo lingüístico, bem como alguns ramos da lingüística quantitativa de textos.

[2] No nível das instâncias, esta contextualização é complementada pelo real contexto do recebimento/produção bem como pelo contexto cognitivo do receptor/produtor.

[3] O exposto a seguir reflete o ponto de vista do autor e não busca reproduzir os escritos já estabelecidos nem determinar suas definições.

[4] Neste sentido, simulações computacionais de processos cognitivos são não-inteligentes; elas não compreendem um comportamento inteligente.

[5] Claramente esta 'formalização' despreza muitas características constitutivas de simulações, realizações e emulações. Assim, isto serve somente como uma 'metáfora formal' para a distinção destes termos.

[6] É exatamente o tipo de modelo simulativo 'autônomo' que será significado a seguir quando utizarmos o termo simulação, abandonando a ampla definição dada acima.

[7] Este conceito parte necessariamente do paradigma da categorização, onde o objetivo da aprendizagem é fixado por meio de um conjunto pré-estabelecido de categorias estáticas (freqüentemente não estruturadas) cujas características quantitativas têm que ser aprendidas por meio de conjuntos de testes pré-categorizados. Para separar modelos de SC deste paradigma deve-se dizer que eles exploram estas categorias como entidades dinâmicas, estruturadas.

Referências

ANDERSEN, P. B. 1990. *A Theory of Computer Semiotics:* Semiotic Approaches to Construction and Assessment of Computer Systems. Cambridge University Press.

__. 2000. Genres as Self-Organising Systems'. Em: *Downward Causation. Minds, Bodies and Matter.* P. B. Andersen, C. Emmeche, N. O. Finnemann, e P. V. Christiansen (eds.). Aarhus. pp. 214-260.

CLARKE, R. 2001. Studies in Organisational Semiotics: an Introduction. *Australian Journal of Information Systems* 8(2): 75-82.

DÖBEN-HENISCH, G. 2002. Alan Matthew Turing, the Turing Machine, and the Concept of Sign. Em: *Das Europ¨aische Erbe der Semiotik; The European Heritage of Semiotics.* W. Schmitz and T. A. Sebeok (eds.). W.e.b. thelem.

HALLIDAY, M. A. K. 1977. Text as Semiotic Choice in Social Context. Em: *Grammars and Descriptions.* T.A. van Dijk e J.S. Petofi (eds.). de Gruyter. pp. 176-225.

HJELMSLEV, L. 1969. *Prolegomena to a Theory of Language*. University of Wisconsin Press.

MARR, D. 1982. *Vision:* a computational investigation into the human representation and processing of visual information. Freeman.

MARTIN, J. R. 1992. *English Text. System and Structure*. John Benjamins.

MEHLER, A. 2002. Quantitative Methoden. Em: *Texttechnologie. Perspektiven und Anwendungen*. H.L. e L. Lemnitzer (ed.). Stau enburg.

PATTEE, H. H. 1988. Simulations, Realizations, and Theories of Life. Em: *Artificial Life. SFI Studies in the Sciences of Complexity*. C.G. Langton (ed.). Addison-Wesley. pp. 63-77.

PEIRCE, C. S. 1993, *Semiotische Schriften* 1906-1913, Vol. 3. Suhrkamp.

RIEGER, B. 2001. Computing Granular Word Meanings. A Fuzzy Linguistic Approach in Computational Semiotics. Em: *Computing with Words*. P. Wang (ed.). Wiley. pp. 147-208.

SEARLE, J. R. 1980. Minds, Brains, and Programs. *The Behavioral and Brain Sciences* 3: 417-457.

STROHNER, H. 1995. Kognitive Systeme. *Eine Einf¨uhrung in die Kognitionswissenschaft*. Westdeutscher Verlag.

STUBBS, M. 1996. Text and Corpus Analysis. *Computer-Assisted Studies of Language and Culture*. Blackwell.

VARELA, F. 1993. *Kognitionswissenschaft, Kognitionstechnik:* eine Skizze aktueller Perspektiven. Suhrkamp.

COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA



CAPÍTULO 6

MÁQUINAS SEMIÓTICAS

Winfried Nöth

MÁQUINA SIMBÓLICAS E SEMIÓTICAS

Máquina simbólica tornou-se uma designação metafórica comum para o computador, mas os semioticistas têm razão ao generalizar esta designação para *máquina semiótica*. Mas o que é uma máquina semiótica? Se é apenas uma máquina envolvida em processos sígnicos, então uma máquina de escrever pode ser chamada de máquina semiótica. Mas se é uma máquina que envolve não apenas processos sígnicos, mas também a capacidade de criar processos de produção e interpretação de signos (ou *semiosis*) então pode haver dúvidas se meros computadores podem ser chamados de máquinas semióticas.

Máquinas simbólicas

Os cientistas da computação chegaram à conclusão, nos anos 50, que os computadores eram mais do que simples máquinas de calcular; eles deviam ser concebidos como *máquinas de processamento*

simbólico (Newell 1980: 137, Nake 1998: 463). Allan Newell (1980) introduziu a noção de sistema de símbolos físicos (*physical symbol system*) para caracterizar sistemas não apenas capazes de processar números, mas também *símbolos*. Com sua teoria de sistemas de símbolos físicos, Newell pretendia construir uma ponte teórica entre a ciência dos seres vivos inteligentes, ou ciência cognitiva, e a ciência das máquinas inteligentes, ou ciência da computação, e a Inteligência Artificial (IA).

Em um sentido bastante distinto, Sybille Krämer (1988) introduziu a teoria de *máquinas simbólicas*. De acordo com a definição de Krämer, uma máquina simbólica é um dispositivo que existe apenas simbolicamente no papel, não tendo qualquer incorporação física real. Tal máquina, em um sentido metafórico, não faz coisa alguma exceto transformar 'seqüências de símbolos'. Um exemplo de tal *máquina* é o algoritmo para multiplicação de números em notação decimal. Um computador, de acordo com esta definição, não é uma máquina simbólica de fato, mas uma espécie de metamáquina, 'uma máquina capaz de imitar qualquer máquina simbólica' (*ibid*.: 2-3).

Este capítulo não vai tratar de máquinas em sentido metafórico mas de máquinas reais de processamento simbólico, como as descritas por Newell. Observe, entretanto, que a definição matemática do conceito de 'máquina' é aplicável a ambas. Uma máquina é um dispositivo que 'determina uma função de suas entradas para as suas saídas' (Newell 1990: 65).

Processamento de signos em computadores

Do ponto de vista da semiótica geral, a mudança histórica das máquinas que processavam apenas números para as que processavam símbolos não constituiu um marco histórico, como sugeriu Newell. Números não são mais do que uma classe de símbolos e a operação com números não é radicalmente distinta da operação com outros símbolos, como Peirce afirma: 'Embora nem todo raciocínio seja computação, é certamente verdadeiro que computação numérica é um raciocínio' (CP 2.56).

Além do mais, computadores não operam apenas com símbolos, mas também com signos icônicos e indexicais (veja seção 2). De acordo com Peirce, um símbolo é um signo que se relaciona com o objeto por ele designado de acordo com 'uma lei ou regularidade'

(CP 2.293). Ambos, palavras e números, pertencem à subcategoria de símbolos remáticos. A maioria dos programas de computador para processamento de texto possui um dicionário que oferece sinônimos para melhorias de estilo. Quando o usuário faz uso dele, o computador produz símbolos remáticos. Máquinas capazes de produção simbólica, neste sentido, são conhecidas desde as primeiras máquinas simbólicas de W. Stanley Jevons e Charles Babbage, no século XIX. Estas eram máquinas lógicas: após a entrada das premissas, o usuário, ao pressionar uma alavanca, obtinha a conclusão como uma resposta automática (Peirce 1887, Ketner 1988, Krämer 1988: 128). Tais máquinas eram não apenas capazes de produzir símbolos remáticos, mas símbolos da categoria do argumento (Nöth 2000a: 67).

Signos indexicais, que dirigem a atenção do intérprete para seu objeto, por meio de uma conexão espacial, causal ou temporal, imediata, são evidentes na programação de computadores e sistemas de processamento de texto, quando o usuário é instruído por meio de setas, cursores, ou por comandos como *atribuir*, *fazer*, *sair se*, ou *continuar se* (Newell 1980: 144-145). Signos icônicos, baseados numa relação de similaridade com seus objetos, também ocorrem em sistemas de processamento de texto. *Copiar e Colar* está entre as operações mais comuns capazes de produzir signos icônicos. O mapeamento, modelagem e até a simulação da realidade pertencem às formas mais complexas de representação icônicas das quais computadores são capazes.

Máquinas semióticas e a semiose das máquinas

A partir daqui, estaremos preocupados com o computador não apenas como *máquina simbólica*, mas também como *máquina semiótica* (Nake 1997: 32), uma máquina que não está restrita ao processamento de símbolos, mas está envolvida em diversos processos sígnicos. Nosso tópico é a *semiose das máquinas*, tal como definido por Andersen (et al. 1997: 548): 'processos sígnicos dentro de máquinas e entre máquinas'.

Entretanto, antes de adotarmos termos como *semiose das máquinas* e *máquina semiótica*, temos de definir a natureza da semiose e do processamento sígnico em geral, e temos de fazer distinções entre diferentes tipos de processos sígnicos, nos quais as máquinas estão envolvidas. Por exemplo, a mediação de signos *por meio* de

máquinas devem ser distinguidas da natureza do processamento sígnico *dentro* das máquinas.

O campo semiótico dos processos sígnicos, de dispositivos técnicos até sistemas vivos, tem sido freqüentemente analisado em termos de dualismos: 'ferramentas vs. instrumentos', 'instrumentos vs. máquinas' e sobretudo 'máquinas vs. seres vivos'. Ao invés de confirmar tais dualismos, tentaremos descrever este campo, dos sistemas semióticos mais simples aos mais complexos, como um *continuum* de processamentos sígnicos, dos mais simples aos mais complexos. Entre os sistemas menos complexos estão aqueles mediados por instrumentos ou dispositivos técnicos como um termômetro, um relógio solar, um termostato ou um sistema de sinalização de trânsito automático. Os sistemas semióticos mais complexos acontecem em seres vivos.

SIGNOS E SEMIOSE, QUASI-SIGNOS E QUASI-SEMIOSE

Existem muitos modelos e definições de signo. Neste capítulo, seguiremos a semiótica de C.S.Peirce (Nöth 2000a: 62-64, 227). Um signo é um fenômeno material, ou meramente mental, relacionado com um fenômeno anterior, objeto do signo, resultando em um outro signo, o interpretante, que fornece uma interpretação do primeiro signo em relação a seu objeto. A semiose é um processo dinâmico no qual o signo, afetado por seu objeto, desenvolve seu efeito no interpretante. O signo não serve como mero instrumento do pensamento, ele tem uma dinâmica própria que é independente de uma mente individual. Além do mais, a semiose não está restrita à produção e interpretação de signos em humanos.

Peirce defende uma tese de continuidade entre 'mente e matéria' que ele chama de sinequismo (CP 7.565; CP 1.172). Esta tese implica na existência de semiose na matéria, em máquinas e mentes humanas?

O paradoxo da máquina semiótica

Se definirmos a semiótica Peirceana como 'a doutrina da natureza essencial das variedades fundamentais das semioses possíveis' (CP 5.488) e *semiose* como ação 'inteligente ou ação triádica do signo' (CP 5.472-73) envolvendo 'uma cooperação entre três sujeitos, como um signo, um objeto e seu interpretante' (CP 5.484), e se aceitar-

mos a 'suposição provisória de que o interpretante é [...] um análogo suficientemente próximo de uma modificação na consciência' (CP 5.485), a idéia de *máquina semiótica* deve parecer uma contradição. A semiótica, de acordo com tais premissas, parece pressupor organismos vivos como produtores e interpretadores de signos. Se a 'ação do signo' pode também se desenvolver em máquinas, se a semiose pressupõe vida, este é um problema a ser examinado tendo como base a semiótica de Peirce.

Sem dúvida, máquinas estão envolvidas em processo sígnicos. Com capacidade para processamento de dados, o computador é certamente uma máquina operando com signos. Mas muitas outras máquinas estão envolvidas em processos sígnicos. Máquinas de escrever, fotocopiadoras, câmeras e gravadores são máquinas que produzem signos. Seriam máquinas semióticas? Uma fotocopiadora não pode ser chamada de máquina semiótica, embora se possa afirmar que ela produz signos. Uma caneta também está envolvida com a produção de signos, mas dificilmente pode ser considerada como causa suficiente de um interpretante.

Apesar de seus critérios de semiose, que sugerem vida como um pré-requisito, Peirce (1887), que freqüentemente usava o termo 'lógico' como um sinônimo de 'semiótico', formulou uma teoria para 'máquinas lógicas' (sem chamá-las de 'máquinas semióticas') muito antes da invenção da Inteligência Artificial (Ketner 1988; Skagestad 1993, 1999; Tiercelin 1993). Mais de um século atrás, ele discutiu as 'máquinas lógicas', inventadas por Jevons e Marquand, e concluiu que estes dispositivos, bem como as máquinas de calcular de seu tempo, eram 'máquinas de raciocínio'. Uma vez que o raciocínio parece ser um processo semiótico, poderíamos concluir que estas máquinas eram máquinas semióticas. Entretanto, Peirce sugere que elas não são, quando conclui que '*toda* máquina é uma máquina de raciocínio' (ibid.: 168). Então, será possível raciocínio sem semiose? Em outro lugar Peirce fornece uma resposta: uma máquina, tal como o tear de Jacquard, embora capaz de raciocinar de acordo com as premissas anteriores, não é capaz da 'produção triádica do interpretante' e opera apenas como um *quasi-signo* (CP 5.473).

Processamento mecânico de signos como quasi-semiose

O termo quasi-signo sugere uma resposta à questão se pode existir semiose em uma máquina de um tipo conhecido por Peirce. Um quasi-signo é similar a um signo apenas em alguns aspectos, mas não pode cumprir todos os requisitos da semiose. Enquanto alguns critérios de semiose podem estar presentes em máquinas, outros estão ausentes. O conceito de quasi-signo sugere então a existência de graus de semioticidade. Quasi-semiose não começa apenas com máquinas de calcular. Ela pode ser encontrada em processos nos quais instrumentos muito mais simples estão envolvidos.

Entre os instrumentos a que Peirce atribui uma função quasi-semiótica está o termostato 'dinamicamente conectado ao aparato de aquecimento e resfriamento, de forma a verificar ambos os efeitos.' A indicação automática de temperatura que ocorre no termostato é apenas uma instância da 'regulação automática' e não cria um interpretante como seu 'significado de saída', Peirce argumenta (CP 5.473). Não existe índice genuíno, mas apenas um quasi-índice, nenhuma semiose, mas uma quasi-semiose.

Quasi-semiose, no termostato, é apenas a redução ('degeneração' é o termo usado por Peirce) de um processo triádico envolvendo um signo (representamen) afetado por um objeto e criando um interpretante para um processo diádico com apenas o signo sendo afetado por seu objeto. A diferença entre os dois tipos de processos é manifesta quando Peirce compara a 'quasi-interpretação' mecânica da temperatura indicada pelo termostato com uma interpretação mental da temperatura indicada por um termômetro.

> A aceleração do pulso é provavelmente um *sintoma* de febre e a elevação da coluna de mercúrio em um termômetro [...] é um *índice* de um aumento da temperatura atmosférica, que, todavia, age nele de forma puramente bruta e diádica. Nestes casos, entretanto, uma representação mental do índice é produzida, e é chamada de *objeto imediato* do signo; e este objeto produz triadicamente o efeito desejado, ou adequado, na forma de outro signo mental (CP 5.473).

Então, quando uma máquina reage de forma causal à temperatura indicada pelo termostato, ela não a interpreta. Neste caso, não há semiose genuína, mas o sinal indicando a temperatura pela qual é afetada de forma causal funciona como um quasi-índice, e a reação mecânica da máquina produzida por este quasi-índice é um

processo de quasi-semiose. Causa e efeito constituem uma relação diádica. A semiose começa a acontecer apenas quando um interpretante é criado para interpretar, por conta própria, esta díada de causa e efeito.

Processamento sígnico em computadores como uma quasi-semiose

A evidência da natureza quasi-semiótica do processamento de dados está na natureza diádica dos signos envolvidos. A visão de que o processamento sígnico em computadores baseia-se em relações diádicas está implícita em uma teoria bastante difundida que estabelece que computadores podem apenas processar *sinais* (Nake 1997: 33), como estímulos mecânicos seguidos por reações automáticas. Winograd e Flores (1986: 86-87), por exemplo, se referem assim ao processamento de sinais: 'as operações de um computador digital podem ser descritas como uma mera sequência de impulsos elétricos que viajam em uma rede complexa de componentes eletrônicos. Estes impulsos não são símbolos de nada.'

Considere os três exemplos de processamento icônico, indexical e simbólico, discutidos anteriormente: 'copiar e colar,' 'sair-se' ou 'dar-sinônimo-de'. Os processos envolvidos claramente constituem relações diádicas entre signos, dentro do computador. De fato, quando Newell (1990: 74-75) descreve processamento de símbolos dentro do computador como um processo relacionando dois símbolos físicos, X e Y, onde X permite o 'acesso à estrutura Y distante', que é transportada por *recuperação* da locação distante para o local', ele oferece uma boa explicação de processos diádicos de quasi-semiose. O que falta para estes signos se desenvolverem, de diádicos para triádicos, é uma relação com o objeto. Relações diádicas são meras relações sem denotação, sem qualquer 'janela para o mundo' relacionando o signo ao objeto da experiência (Nöth 1997: 209-210). Concluímos, assim, que signos icônicos, indiciais e simbólicos, com os quais o computador opera, são quasi-signos.

Semiose na interface entre humanos e computadores

Apesar dos processos semióticos dentro de máquinas serem quasi-semióticos, processos nos quais máquinas servem como mediadores, na semiose humana, são certamente processos de semiose genuína. Se um sinal de trânsito é um signo genuíno para um moto-

rista, então um semáforo não é um signo menos genuíno. Neste sentido, o processamento sígnico na interface entre humanos e computadores é semiose genuína. Signos são produzidos por humanos, mediados por máquinas, e interpretados por humanos. Nessa cadeia clássica de comunicação, o computador pertence à mensagem. Remetente e destinatário humanos são, ou duas pessoas distintas, ou a mesma pessoa em uma situação de auto-comunicação. Nestes processos de comunicação mediados por computador, ele serve como uma extensão semiótica da semiose humana; ele é usado como a mais poderosa ferramenta para manipulação mais eficiente da semiose humana. Trata-se do desenvolvimento mais recente na extensão semiótica de humanos em um processo cultural que começou com a invenção da pintura, escrita, impressão, fonografias, máquinas de escrever e outras mídias (cf. Popper 1972: 238-39). Entretanto, as mensagens produzidas por um computador, na interface de humanos e máquinas, são, ou mensagens transportadas por um remetente humano e mediadas pelo computador, ou são quasi-signos resultando de uma extensão automática e determinística da semiose humana.

MÁQUINAS COM MENTE VS. MENTES MECÂNICAS

Todavia, deve-se determinar se um computador também pode ser um agente em um processo semiótico genuíno. Pode ele ser a fonte de uma 'ação triádica do signo, ou inteligente', por si só? Talvez o processamento de signos em computadores esteja apenas no nível mais rudimentar, de redução à sinalização eletrônica, e portanto de quasi-semiose. Talvez a complexidade da semiose do computador seja insuficientemente descrita neste nível, como o cérebro, quando suas operações são descritas como seqüência de sinais que ocorrem como entrada e saída de bilhões de neurônios.

A questão sobre se é possível a semiose em computadores está intimamente relacionada a questões como: Computadores podem pensar? Teriam intenções? Teriam mente? Antes de tratar da teoria da mente de Peirce, e de suas considerações sobre a possibilidade de semiose genuína em máquinas, introduziremos um argumento clássico contra o comportamento mental em computadores, que será constrastado com o argumento de que a máquina pode realizar atividades mentais.

Agentes não-mentais na sala chinesa de Searle

A visão do computador como uma mera máquina de processamento de sinais tem sido defendida por John Searle (1980) em categorias mentalistas. O núcleo do argumento é o seguinte: um computador trabalhando de acordo com um algoritmo não pode ser uma máquina mental, uma vez que não pode entender os símbolos com os quais opera. Searle explica seu argumento por meio da famosa parábola da sala chinesa, onde mensagens são processadas por pessoas que não entendem o significado das palavras. As pessoas nesta sala são americanas, falantes do inglês, e recebem mensagens em chinês. Todavia elas são capazes de processá-las com base em instruções numéricas que informam como combinar e correlacionar os elementos das mensagens. Conseqüentemente, os falantes (o computador) não entendem (e portanto não são afetados pela semiose)

> porque as manipulações formais de símbolos por eles próprios não têm nenhuma intencionalidade; elas são, na verdade, sem sentido; elas não são nem sequer manipulações de *símbolos*, uma vez que os símbolos não simbolizam nada. [...] A intencionalidade da forma como computadores parecem apresentar está tão somente na mente daqueles que os programaram, aqueles que enviaram as mensagens e aqueles que interpretaram a saída (Searle 1980: 422).

Por meio desta parábola dos agentes 'cegos' trabalhando mecanicamente dentro da máquina sem mente, Searle acredita ter dado um choque fatal no mito do computador como máquina mental. Entretanto, seu argumento sofre de uma influência Cartesiana onde é possível dividir, de forma cristalina, os trabalhos mental e mecânico. Seu argumento não é realmente válido contra a idéia de mente no computador. Além do mais, para realização de seus trabalhos mecânicos, os pobres americanos na sala chinesa precisam ter mentes e intenções. Assim, o trabalho que eles fazem deve ser mental, e a máquina, da qual eles são uma metáfora, tem de ser uma máquina mental.

A mente no trabalho manual, mecânico e mental.

Do ponto de vista da história cultural, *máquina* é definido como um aparato que requer uma entrada de força ou energia para realizar certas tarefas que substituem, e portanto economizam, o trabalho de humanos ou animais. Um carro requer a entrada de gasolina e economiza o trabalho de humanos e cavalos. Uma má-

quina de lavar requer a entrada de eletricidade e economiza o trabalho manual de lavar à mão.

Seguindo este raciocínio, o computador é uma máquina que economiza trabalho mental (Nake 1992: 185, Santaella 1998: 124). Contrastando com as generalizações anteriores, de que máquinas puramente mecânicas servem para substituir apenas trabalho manual ou muscular, o computador deve ser então uma máquina mental, uma vez que serve como substituto para o trabalho mental. Entretanto, onde termina o trabalho manual e começa o mental? Pode esta questão ser respondida sem influência Cartesiana?

Vamos considerar inicialmente o trabalho manual e os vários dispositivos culturais que foram inventados para substituí-lo. De fato, a economia de trabalho começa com dispositivos que foram inventados muito tempo antes da primeira máquina. O simples instrumento de escrita de uma caneta-tinteiro, por exemplo, é um instrumento que economiza trabalho, uma vez que o uso de seus predecessores, como a pena, requeriam o trabalho de usar um tinteiro separamente durante a escrita. É claro que a canela-tinteiro não requer entrada de energia, não sendo portanto uma *máquina* de escrita, mas apenas uma *ferramenta* de escrita.

Será uma máquina de escrever mecânica, como a equivalente em alemão *Schreibmaschine* sugere, uma 'máquina de escrita', ou uma mera ferramenta? Dado que nas antigas máquinas de escrever não existe entrada de energia, nem uma economia real de energia muscular em comparação à escrita manual, uma máquina de escrever dificilmente é mais do que uma simples ferramenta de escrita. Uma máquina de escrever elétrica, por outro lado, é certamente uma máquina. Ela requer eletricidade como entrada e facilita o trabalho manual reduzindo o esforço muscular. Será que ela também economiza trabalho mental, como o computador, ou apenas trabalho muscular?

Se não existe grande diferença entre a escrita à máquina e a escrita à mão, no que se refere ao esforço manual investido na tarefa de escrita, por que as máquinas de escrever foram inventadas afinal de contas? Aparentemente, a escrita à máquina não produz facilidade de escrita, mas facilidade de leitura, devido aos caracteres padronizados e regulares, linhas e parágrafos. Maior facilidade de leitura, entretanto, também significa economia de trabalho mental. Assim, a máquina de escrever mecânica, muito

antes do computador, já era uma máquina que servia para economizar trabalho mental.

Outra máquina que sem dúvida facilita o trabalho mental é a máquina de calcular. Calcular é um trabalho mental e uma máquina que calcula para seu usuário é uma máquina que economiza trabalho mental. Por outro lado, é verdade que, usualmente, sem uma máquina de calcular, nós calculamos as tarefas mais complexas por meio de operações manuais, escrevendo os números em ordem e linha, de forma a quebrar a tarefa complexa em operações elementares mais simples. Isto torna o cálculo um trabalho manual, de forma que calculando com uma máquina de calcular não apenas economizamos trabalho mental mas também trabalho manual.

Uma máquina como a máquina de costura parece ser uma das candidatas menos prováveis a máquina semiótica, devido ao fato de ter sido inventada exclusivamente para o propósito de economia de trabalho manual. Entretanto, não é o tipo de trabalho que ela economiza, notadamente a costura à mão, também um tipo de trabalho mental? Afinal de contas, o corte e manipulação do tecido, a agulha e o alfinete requerem planejamento cuidadoso e coordenação dos movimentos. É necessário pensar, antes e durante a operação de costura.

Resumindo, a distinção entre trabalho manual e trabalho mental não é clara. Todas as máquinas economizam trabalho mental e manual. Não é por acidente que as áreas do cortex humano que coordenam nossas operações manuais são tipicamente grandes. A área cerebral que coordena os movimentos das mãos e braços humanos não é menor do que aquela que coordena as expressões faciais e os movimentos da língua e do maxilar durante a articulação da fala (Geschwind 1982: 112), e isto não é realmente surpreendente se considerarmos os paralelos evolutivos entre as atividades manuais e comunicativas (Leroi-Gourhan 1964-65: 188-89). Agora, se todas as máquinas economizam trabalho mental e são portanto máquinas com mentes, qual é a diferença entre mentes mecânicas e humanas?

Máquinas de raciocínio e mentes mecânicas

É diferente a resposta de Peirce à questão da mente na máquina. Apesar de sua teoria de quasi-semiose mecânica, seu argumento é que enquanto máquinas não funcionam como mentes humanas em

todos os aspectos, elas o fazem em parte; ao mesmo tempo máquinas devem ser vistas como mentes mecânicas.

Adicionalmente à sua teoria de quasi-semiose em máquinas, que destaca a diferença entre a semiose humana e o processamento sígnico em máquinas, Peirce, em sua teoria das máquinas lógicas, também considerou as similaridades entre humanos e máquinas (Ketner 1988, Tiercelin 1993: 228ff). Ao contrário de Searle, Peirce argumenta que a mente humana funciona como uma máquina, em certos aspectos. Este argumento soa reducionista, mas ele certamente não afirma que a mente humana é uma máquina. Apenas quando resolve uma tarefa que uma máquina lógica ou máquina de calcular *também* pode resolver, isto é, pela execução de regras de um algoritmo de forma quase mecânica, a mente humana funciona como uma máquina:

> Insisto no fato que, de forma similar, um homem possa ser considerado uma máquina que converta, digamos, uma sentença escrita expressando uma conclusão, tendo ele, o homem-máquina, sido alimentado com uma afirmação escrita de algum fato, como uma premissa. Desde que essa atuação não seja mais do que aquela que uma máquina faria, não tem nenhuma relação essencial com a circunstância de que a máquina funciona por engrenagens, enquanto um homem funciona por um arranjo não muito bem conhecido de células cerebrais (CP 2.59).

De acordo com esta teoria sinequística de transição gradual entre matéria e mente, Peirce não apenas conclui que a mente humana, quando resolvendo um problema matemático ou lógico, como uma máquina mental, mas também que as máquinas de calcular e as máquinas lógicas de seu tempo, eram 'máquinas de raciocínio.' Esta similaridade entre o pensamento humano e o mero 'raciocínio' mecânico, de acordo com Peirce, decorre da herança evolutiva comum da natureza biológica e física: ambos, o cérebro humano e as leis da mecânica, estão sob as mesmas restrições cosmológicas de forma que um certo grau de similaridade entre os dois pode ser assegurado (cf. Nöth 2001a, 2002). O modo de processamento sígnico comum a humanos e máquinas é a iconicidade diagramática:

> O segredo de todas as máquinas de raciocínio, no fundo, é muito simples. É o de que qualquer relação entre os objetos sobre os quais se raciocina está destinada a ser o ponto focal do raciocínio puro; esta mesma relação geral deve poder ser introduzida entre certas partes da máquina (Peirce 1887: 168).

Sobre esse assunto, entretanto, não apenas uma máquina lógica, mas

> toda máquina é uma máquina de raciocinar, na medida em que existem certas relações entre suas partes, relações que envolvem outras relações não explicitamente expressas. Uma peça do aparato para realizar um experimento físico ou químico é também uma máquina de raciocínio, com a diferença de que ela não depende das leis da mente humana, mas da razão objetiva incorporada nas leis da natureza. Conseqüentemente, não é figura de linguagem dizer que o alambique e o cucurbis do químico são instrumentos do pensamento ou máquinas lógicas (ibid.).

Quasi-mente no tinteiro

Se não apenas máquinas lógicas, mas também todas as outras máquinas, e até mesmo instrumentos técnicos, são instrumentos do pensamento aprimorados com a capacidade de raciocínio, então devemos concluir que máquinas provocam mentes. De fato, Peirce chega ao ponto de atribuir mente e pensamento ao mundo físico: 'O pensamento não está necessariamente conectado com um cérebro. Ele aparece no trabalho das abelhas, nos cristais e por todo o mundo puramente físico' (CP 4.551). A teoria semiótica da mente que fundamenta esta afirmação está além do escopo deste capítulo (ver Santaella 1994). Podemos indicar alguns de seus aspectos, no nosso estudo do enigma da mente na máquina. Neste contexto, é relevante destacar que Peirce, ao falar de 'pensamento não humano' (CP 4.551) introduz o conceito de *quasi-mente*, para fazer uma distinção entre a mente, no sentido da psicologia cognitiva, e processos de semiose associados com signos 'num sentido muito amplo' (ibid.).

Assim, quasi-semiose e quasi-mente são o que encontramos nas 'máquinas mentais' e 'mentes mecânicas'. Peirce também desenvolve o argumento de que, em um sentido mais amplo, a mente está localizada não apenas no cérebro de um escritor, mas também na materialidade de seu meio semiótico, isto é, na tinta.

> Um psicólogo remove o lóbulo de meu cérebro [...] e então, quando descubro que não posso mais me expressar ele diz, 'veja que sua faculdade da linguagem estava localizada naquele lóbulo'. Sem dúvida que estava; e assim, se ele tivesse roubado meu tinteiro, eu não seria capaz de continuar minha discussão até que conseguisse outro. Sim, os próprios pensamentos não viriam a mim. Então minha faculdade de

> discussão está igualmente localizada no meu tinteiro. É uma localização no sentido em que uma coisa pode estar em dois lugares ao mesmo tempo (CP 7.366).

A interpretação desta passagem enigmática de 1902 tem várias facetas (Skagestad 1993, 1999; Tiercelin 1993: 240), mas em nosso contexto o argumento de Peirce é especialmente relevante, visto que devemos procurar a mente 'em dois lugares ao mesmo tempo', em seu cérebro, local interno de produção de signos, e no tinteiro, local da materialização externa do signo. Ambos representam aspectos inseparavalmente unidos da semiose, como dois lados de uma moeda. Justificativas deste argumento, no que refere à unidade essencial das manifestações internas e externas do signo, podem ser encontradas no pragmaticismo de Peirce. Ele provê duas chaves para o entendimento do enigma da mente no tinteiro: teoria da unidade do signo, com sua representação externa, e a teoria do pensamento e da ação.

A teoria da unidade do signo e sua representação estabelece que 'pensamento e expressão são realmente um' (CP 1.439). Pensamento, no sentido de um traço de memória cerebral, e sua expressão na forma de manifestação escrita, são os dois lados do mesmo signo. A palavra escrita não é meramente um instrumento externo produzido por um cérebro humano, e usado por um ser humano, para um propósito externo específico, como a teoria instrumental do signo assevera (ver Nöth 2000a). Contra a visão instrumental do signo, Peirce defende que a idéia, ou pensamento, transportado pelo signo, não pode existir *antes* deste signo ser manifesto externamente; ao invés disso, existem, simultaneamente, a idéia e sua representação. Nem o significado, no sentido do interpretante, precede o signo, já que ele é o efeito, e não a causa do signo. Se o pensamento não precede sua representação, mas existe semioticamente com ela, a busca pelo pensamento e pelo significado na 'caixa do cérebro' seria uma busca em vão, porque há uma manifestação externa que testemunha a natureza deste pensamento. Uma vez que idéias representadas por palavras, textos ou livros não precedem tais manifestações, a conclusão de Peirce é que o signo não pode ser localizado no cérebro, mas precisa ser investigado nos signos que resultam da atividade cerebral. Atento ao segundo lado da moeda semiótica, Peirce conclui que 'é bem verdade que os pensamentos de um escritor vivo estão em qualquer cópia impressa de seu livro, mais do que em seu cérebro' (CP 7.364). Em um contexto diferente, em que o tópico é o estilo do escritor

('boa linguagem'), Peirce expressa sua idéia da unidade do signo e pensamento da seguinte forma: 'É incorreto afirmar que uma boa linguagem é importante para um bom pensamento, porque aquela é a essência deste' (CP 2.220).

O princípio da unidade do pensamento e da ação fornece outra chave para o enigma da mente no tinteiro. A mente de um autor não pode ser reduzida àquilo que acontece no *cérebro*, uma vez que o processo de escrita também envolve a atividade manual externa de usar o meio da tinta para produzir a palavra escrita. 'Meu lápis é mais inteligente que eu,' Einstein costumava afirmar com referência ao cálculo manual no papel (cf. Skagestad 1993: 164). O cálculo escrito não é apenas uma alternativa semiótica ao cálculo falado e mental, mas uma operação que permite o desenvolvimento de argumentos mais difíceis e a solução de problemas mais complexos, visto que a fixação dos signos no papel tem a vantagem de aumentar nossa memória. Este efeito de externalização de nossa memória é uma das razões pelas quais os pensamentos vêm ao escritor durante a escrita. Além do mais, os pensamentos que vêm a nós não são os mesmos daqueles quando pensamos no mesmo assunto. Esta diferença está manifesta na distinção entre o estilo oral e o escrito. Hoje, depois da tese de McLuhan sobre 'o meio é a mensagem', nós podemos também presumir que os pensamentos que vêm a nós quando escrevemos, por meio de uma máquina, não são iguais, em todos os aspectos, àqueles baseados na caneta como meio.

A conclusão desta linha de argumento é que, por um lado, existe uma (quasi)-mente não apenas no cérebro, mas também na máquina, e, por outro, que ela é apenas uma condição necessária, mas ainda insuficiente de semiose genuína. As condições restantes serão o tópico da seção final deste capítulo.

Controle, auto-controle e autopoiese

Apesar de serem capazes de raciocinar, as máquinas lógicas do século XIX ainda não possuíam a capacidade de semiose genuína que Peirce costumava associar a *auto-controle*. Uma máquina não tem auto-controle se é completamente controlada por suas entradas. Seriam todas as máquinas deste tipo, ou existem máquinas que começam a exercer controle sobre si mesmas?

Controle

De acordo com Pattee (1997: 29), não apenas auto-controle, mas também controle ambiental é uma característica distintiva de organismos biológicos:

> Controles são lógicos e condicionais. A vida se originou com controles semióticos. Controles semióticos requerem medida, memória e seleção, nenhum dos quais pode ser totalmente descrito por leis físicas que, ao contrário de sistemas semióticos, são baseados em energia, tempo e taxas de mudança. [...] Para funcionar eficientemente, controles semióticos, em todos os níveis, devem prover descrições simples do comportamento dinâmico complexo dos sistemas de entrada e saída, que chamamos de sensores, detectores de características, reconhecedores de padrões, dispositivos de medida, transdutores, construtores e atuadores.

Máquinas podem também exercer controle ambiental. Um simples termostato, e mecanismos de realimentação, servem a propósitos de controle ambiental. Mas ao exercer tal controle, a maioria deles é apenas uma extensão dos humanos. Enquanto uma máquina pode ter controle sobre seu ambiente, é o usuário humano que detém o controle da máquina.

Máquinas determinísticas como sistemas alopoiéticos

Em um manuscrito de 1906, Peirce descreve assim a ausência de autocontrole na quasi-semiose mecânica: 'Ninguém provou que um motor automático não possa exibir auto-controle, além de um autoajuste específico para o qual tenha sido construído; mas ninguém jamais foi bem-sucedido na construção de tal máquina' (MS 498, Ketner 1988: 43). Em conseqüência, 'toda máquina [...] é destituída de tal originalidade, de qualquer iniciativa. Ela não pode encontrar seus próprios problemas, ela não pode se alimentar. Ela não pode se orientar entre diferentes procedimentos possíveis' (1887: 168). Tais máquinas são estritamente *determinísticas*, como Ketner (ibid.) as chama, máquinas que podem apenas 'fazer tipos especiais de coisas para as quais foram projetadas', como Peirce (1887: 169) acrescenta. O controle, em uma máquina determinística, vem 'de fora', do engenheiro que a projetou e do usuário que a manipula. A máquina não é um agente autônomo.

O critério de autonomia foi descrito como uma das características distintivas da vida, relativamente aos sistemas não-vivos. Em teoria de sistemas, o termo *autopoiese* é usado para descrever um

sistema que evidencia este tipo de autonomia devido ao autocontrole (veja abaixo). Quando o controle vem de um lugar qualquer, de fora, o sistema é um *sistema alopoiético* (Schmidt 1987: 22-23). As máquinas consideradas até então são sistemas alopoiéticos. Entretanto, a diferença entre sistemas auto e alopoiéticos é de grau, e elementos de autopoiese e de alopoiese podem ser encontrados em robôs e criaturas de vida artificial.

Automata, controle e autocontrole

A autonomia das máquinas começa com a intervenção dos autômatas. Em comparação com uma máquina de escrever elétrica, por exemplo, um computador realiza muitas das subtarefas de escrita e produção de textos automaticamente. Em contraste com a escrita a máquina, o processamento digital de textos permite formatação automática, a correção de erros de ortografia e a impressão de todo o texto por meio de simples comandos. Enquanto ambas, as máquinas de escrever e os computadores, são máquinas, apenas o computador realiza tarefas suficientemente automáticas para merecer a designação de *autômato*.

Etimologicamente, 'automático' significa 'por conta própria'. Um autômato, portanto, é um sistema capaz de realizar suas tarefas por conta própria. Entretanto, a capacidade de agir por conta própria sugerida pelo nome 'autômato' não é de forma alguma genuína. Nenhum autômato opera com tanta autonomia quanto um ser vivo. Apenas seres vivos têm um 'si próprio', permitem autocontrole e ação autônoma. Um dos fundamentos semióticos deste tipo de autocontrole, que falta nas máquinas, mas é característico de organismos vivos, é a auto-referência (Nöth 2000b). Um autômato determinístico não tem auto-referência. Ele não é auto-referencial, mas *alo-referencial*, isto é, é um sistema capaz de referencializar apenas o ambiente, e não a si próprio. Auto-referência é uma necessidade biológica para o ser vivo, uma vez que para sobreviver ele precisa ter a capacidade para distinguir entre si mesmo e seu ambiente.

Autopoiese e auto-reprodução

Autopoiese em sistemas vivos significa que o sistema é não apenas capaz de auto-referência e autonomia em relação a seu ambiente, mas que também é capaz de se manter, por conta própria, e, finalmente, de se reproduzir. Máquinas não são autopoiéticas,

mas alopoiéticas, sistemas produzidos e mantidos por humanos. Todavia, a distinção entre sistemas alopoiéticos e autopoiéticos, e mais genericamente entre engenharia e biologia, não é mais tão clara. Por outro lado, dúvidas sobre a autonomia genuína da consciência humana foram levantadas. Freud, por exemplo, diria que humanos não agem como seres autônomos. Outras evidências, de como a autonomia da ação humana e o destino dos humanos, em geral, são determinados por fatores independentes de cada criatura, vêm da biologia evolucionária e da genética contemporânea.

Por outro lado, somos confrontados com programas de computador, autômatos e robôs que não mais parecem meros artefatos alopoiéticos, mas começam a evidenciar características de sistemas autopoiéticos. Vida artificial está sendo criada em telas de computador e a possibilidade de produzir robôs capazes de automanutenção, e até auto-reprodução, está sendo explorada. O biólogo Kawade (1999: 373), por exemplo, chega ao ponto de prever um limiar do ponto de encontro da alopoiese com a autopoiese e, então, do que ele acredita ser o fim da diferença essencial entre sistemas mecânicos e sistemas vivos:

> Se, em um futuro previsto, 'sistemas moleculares auto-reproduzíveis' forem criados pela mão humana, [...] então esta distinção vai também desaparecer. Mesmo que a síntese completa de uma célula viva não seja alcançada, várias estruturas orgânicas artificiais que realizam parte das funções de células completas ou tecidos naturais e organismos serão provavelmente feitas num futuro próximo, tornando vaga a fronteira entre a máquina e a coisa viva.

Em 1948, John von Neumann trabalhou no projeto de um autômato com capacidade de auto-reprodução (Neumann 1966; ver Emmeche 1994: 56). A parte central desta máquina consistia em um dispositivo A com a capacidade de aproveitar 'material cru' do ambiente para produzir, de acordo com as instruções de um duplicador B, e um comando D para um controlador C, como sua saída, um mesmo autômato com os componentes de A, especificamente 'A', 'B', 'C' e 'D'. Tal como Etxeberria e Ibáñez (1999: 295) afirmaram, o processo de auto-reprodução automática neste autômato é um processo semiótico por que a máquina constrói sua cópia de acordo com uma autodescrição interna. O autômato, por assim dizer, pode e deve ler a si próprio para se reproduzir. Auto-reconhecimento e leitura de si próprio pressupõem auto-referên-

cia, de forma que um autômato auto-reproduzível é um sistema auto-referencial.

Apesar da similaridade entre a autopoiese de tal sistema (auto-reproduzível) e um organismo biológico, existe também uma importante diferença. O autômato auto-reproduzível é desprovido do tipo de criatividade genética que acontece na reprodução biológica, e que é a fonte da diversidade das espécies. Uma máquina capaz de produzir uma réplica exata de si mesma é ainda uma máquina determinística, visto que sua saída é precisamente determinada pelo projeto da máquina. Von Neumann chamou este fenômeno de o *limite da complexidade*: 'Quando sistemas artificiais geram objetos, há uma degradação de complexidade entre o agente que constrói e o objeto construído; ao mesmo tempo, sistemas biológicos podem manter, e até mesmo aumentar, o nível de complexidade de seus produtos' (*ibid.*).

Se a ambiciosa máquina auto-reproduzível de von Neumann nunca foi construída, a continuação de seu projeto levou ao desenvolvimento de uma nova geração de *autômatas celulares* auto-reproduzíveis e à atual pesquisa em Vida Artificial, que tem sido bem sucedida na simulação de várias formas de sistemas artificiais com capacidade de auto-organização e auto-reprodução (Cariani 1998; Etxeberria & Ibáñez 1999). Os descendentes desta linha de pesquisa com a qual a maioria de nós está familiarizada são os vírus de computador.

PROPÓSITO, EXPERIÊNCIA E MÁQUINAS GENUINAMENTE SEMIÓTICAS

A distinção entre causalidade determinística, ou eficiente, e final, ou teleológica, é a principal chave para entender as idéias de Peirce sobre semiose (Santaella 1999) e sobre máquinas semióticas. Entretanto, a semiose genuína precisa além disso de criatividade e de habilidade para transformar signos em ação.

Propósito e causalidade final

Em uma passagem que antecede sua argumentação sobre o pensamento no tinteiro, Peirce define assim o seu foco: 'Acredito que propósito, ou melhor, causação final, cujo propósito é a modificação consciente, é o assunto essencial do estudo dos psicólogos; e

que consciência é um acompanhamento especial, e não universal, da mente' (CP 7.366). Há propósito, ou direcionalidade, na ação da semiose por causa do caráter normativo dos signos. Para Pape (1993: 586), 'ao criar e usar signos, nosso ideal é entender e representar o que queremos: nos direcionamos para um resultado que o comportamento do signo quer aproximar.' Embora o uso de signos seja determinado por hábitos, o objetivo do uso do signo só pode ser realizado por aproximação. É por isso que semiose genuína não é mecanicamente determinística, mas abre espaço para auto-correção, criatividade e 'crescimento simbólico' (CP 2.302).

Uma máquina munida de mente, e não apenas de uma quasi-mente, deve então perseguir um propósito semiótico de forma autônoma. Enquanto a causação eficiente, como é característica de máquinas determinísticas, cria 'uma compulsão que age de forma a fazer uma situação começar a mudar de forma perfeitamente determinada' (CP 1.212), causação final, em uma semiose genuína, 'não determina de que forma um particular é obtido, mas apenas que o resultado deve ter um certo caráter geral' (CP 1.211). Existe então causação final quando um signo não é determinado por uma força mecânica, mas por uma norma semiótica ou hábito que não é seguida cegamente, mas que permite certa criatividade na produção e interpretação sígnicas. São tais máquinas genuinamente semióticas possíveis?

Peirce ilustra a diferença entre causação final e eficiente por meio do seguinte exemplo:

> Acerto um tiro na asa de uma águia. Visto que meu propósito — um tipo especial de causa final ou ideal — é acertar o pássaro, eu não atiro diretamente nele, mas um pouco à frente dele, permitindo a mudança de posição durante o tempo em que a bala demorar para percorrer a distância. Este é um caso de causação final. Mas após a bala deixar o rifle, o caso é revertido para a estúpida causação eficiente (CP 1.212).

Enquanto o rifle é uma máquina meramente determinística, o caçador está envolvido uma semiose genuína, perseguindo um objetivo cuja execução requer a operação inteligente de alcançar 'um resultado geral que pode ser obtido a um certo tempo, de uma certa forma, e em outro instante de uma outra forma' (CP 1.211). Quando Peirce traçou estas distinções entre o caçador que planejava e o rifle determinístico, distinguiu semiose humana de uma quasi-semiose mecânica. Hoje sabemos que mísseis automáticos são muito mais capazes de acertar um alvo em movimento do que

qualquer caçador humano. Tal míssil, devemos concluir, é uma máquina semiótica genuína.

Robôs, experiência e pragmática semiótica

É bem conhecido que a distância entre a semiose humana e a semiose em máquinas está decrescendo mais e mais (Cariani 1998). Está além do escopo deste capítulo determinar, de forma mais precisa, o ponto onde a vida artificial e a semiose genuína em máquinas começa. Nem todos os cientistas da computação concordam, por exemplo, que computadores executando programas com Inteligência Artificial são máquinas semióticas genuínas (como Ketner [1988: 56-58] conjectura) ou se são apenas máquinas determinísticas (como Fetzer [1990: 37] argumenta).

Um computador sem uma janela para seu ambiente está apenas envolvido em uma semiose sintática, e talvez semântica, mas não pragmática. De acordo com o princípio da unidade do signo (ou pensamento) e ação (veja 3.4), a dimensão pragmática do processamento sígnico é um critério adicional da semiose completamente desenvolvida. Em uma carta de 1887, Peirce discutiu como sendo uma das diferenças entre o processamento sígnico em máquinas e humanos:

> A lógica formal concentra toda a sua atenção na parte menos importante do raciocínio, uma parte tão mecânica que pode ser realizada por uma máquina, e imagina-se que isso seja tudo o que há no raciocínio. De minha parte, acredito que o raciocínio é a observação de relações, principalmente por meio de diagramas e equivalentes. É um processo vivo. [...] O raciocínio não é feito pelo cérebro sem qualquer ajuda, mas precisa da cooperação dos olhos e das mãos (em Ketner & Stewart 1984: 208-209).

O aprendizado, a partir da experiência com o ambiente, e autocorreção automática, são elementos adicionais essenciais de uma máquina semiótica genuína (Nöth 1997). Um robô que aprende a partir de sua própria experiência em sua orientação no ambiente, e que reage pela reconstrução do projeto de seus próprios programas de forma a melhorar sua eficiência futura, não é mais uma máquina semiótica determinística, mas genuína (Nöth 2001b).

Conclusão e a visão de máquinas poéticas

Um dos resultados deste estudo é que nenhum dos critérios de semiose encontra-se completamente ausente no mundo das máqui-

nas. De um lado, os conceitos de semiose e quasi-semiose de Peirce são tão amplos que nem mesmo matéria e mente estão separados por um limiar semiótico. Por outro lado, a história da engenharia têm trazido tantos tipos de máquinas inteligentes que nenhum critério semiótico pode ser dito como estando completamente ausente do mundo das máquinas. As diferenças que restam entre semiose de máquinas e de humanos são uma questão de grau. Esta diferença é particularmente evidente se considerarmos a característica da criatividade semiótica. Enquanto Peirce acreditava que 'todas as máquinas [...] são destituídas de qualquer originalidade,' (1887: 168), somos hoje confrontados com os primeiros passos na direção da criatividade nas máquinas.

Uma máquina capaz de raciocinar de forma não meramente determinística, mas criativamente, deve ser não apenas capaz de raciocínio dedutivo, mas também abdutivo. Os primeiros passos para transpor este limiar semiótico da quasi-semiose para a semiose genuína foram dados nas pesquisas em Inteligência Artificial (ver Josephson e Josephson, eds. 1994). Um nível ainda mais alto de autopoiese semiótica exigiria a criatividade para produzir não apenas imagens, mas pinturas, não apenas textos, mas textos criativos, novelas e poesias.

Peirce não queria excluir, em princípio, a possibilidade de que uma máquina genuinamente semiótica pudesse um dia ser inventada. Mas ele sabia que a engenharia de seu tempo não tinha avançado além do desenvolvimento de máquinas determinísticas capazes de 'raciocinar' com capacidade bastante limitada. Há um século, quando tais máquinas determinísticas eram ainda operadas de forma manual, a visão de uma semiose genuína lembrou Peirce (1887: 165) a Academia de Lagado, nas Viagens de Gulliver (III.5), de Jonathan Swift. Os membros desta academia possuíam uma máquina semiótica genuína. Ela não era apenas uma máquina capaz de raciocinar, mas também de criar poesia, e 'a pessoa mais ignorante, a um custo razoável, e com pequeno trabalho físico, poderia escrever livros em filosofia, poesia, política, direito, matemática e teologia, sem a menor assistência de algum gênio ou estudo' (Peirce 1887: 165). Enquanto a invenção de tal máquina tem sido um sonho, desde o tempo de Swift, permanece perturbadora a perspectiva de um mundo no qual escravos semióticos poderiam, um dia, não apenas fazer todo trabalho necessário e desnecessário, manual e mental, mas também tornar supérflua a criatividade humana.

Referências

ANDERSEN, PETER BØGH et al. 1997. Machine semiosis. Em: *Semiotik:* Ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur, Bd. 1. Posner, Roland, et al., eds. de Gruyter. pp.548-571.

CARIANI, PETER. 1998. Towards an evolutionary semiotics: The emergence of new sign-functions in organisms and devices. Em: *Evolutionary Systems:* Biological and Epistemological Perspectives on Selection and Self-Organization. van de Vijver, Gertrudis et al. (eds.) Kluwer. pp.359-376.

EMMECHE, CLAUS. 1994. *The Garden in the Machine:* The Emerging Science of Artificial Life. Princeton: Univ. Press. - German transl. Das lebende Spiel. Rowohlt.

ETXEBERRIA, ARANTZA e JESÚS IBÁÑEZ. 1999. Semiotics of the artificial: The 'self' of self-reproducing systems in cellular automata. *Semiotica* 127: 295-320.

FETZER, JAMES H. 1990. *Artificial Intelligence:* Its Scope and Limits. Kluwer.

GESCHWIND, NORMAN. 1982. Specialization of the human brain. Em: *Human Communication:* Language and its Psychobiological Bases. William S-Y Wang, (ed.) (Readings from Scientific American). W. H. Freeman, pp.110-119.

JOSEPHSON, JOHN R.; JOSEPHSON, SUSAN G. (eds.) 1994. *Abductive Inference*. Cambridge Univ. Press.

KAWADE, YOSHIMI. 1999. The two foci of biology: Matter and sign. *Semiotica* 127: 369-384.

KETNER, KENNTEH LAINE. 1988. Peirce and Turing: Comparisons and conjectures. *Semiotica* 68: 33-61.

KETNER, KENNETH LAINE; STEWART, ARTHUR F.. 1984. The early history of computer design: Charles Sanders Peirce and Marquand's logical machines. *The Princeton University Library Chronicle* 45.3: 187-211.

KRÄMER, SYBILLE. 1988. *Symbolische Maschinen*. Wiss. Buchgesellschaft.

LEROI-GOURHAN, ANDRÉ. (1964-65) 1988. *Hand und Wort:* Die Evolution von Technik, Sprache und Kunst. Suhrkamp.

NAKE, FRIEDER. 1992. Informatik und die Maschinisierung von Kopfarbeit. Em: Wolfgang Coy (eds.) *Sichtweisen der Informatik*. Vieweg, pp.181-207.

__. 1997. Der semiotische Charakter der informatischen Gegenstände. *Semiosis* 85-90: 24-35.

__. 1998. Schwierigkeiten beim semiotischen Blick auf die Informationsgesellschaft. Em: *Knowledge Management und Kommunikationssysteme*. Harald W. Zimmermann e Volker Schramm (eds.) Konstanz: Universitätsverlag. pp. 455-468.

NEUMANN, JOHN VON. 1966. *The Theory of Self-Reproducing Automata*, Arthur W. Burks (ed.). Illinois Univ. Press.

NEWELL, ALLEN. 1980. Physical symbol system. *Cognitive Science* 4: 135-183.

__. 1990. *Unified Theories of Cognition*. Harvard Univ. Press.

NÖTH, WINFRIED. 1997. Representation in semiotics and in computer science. *Semiotica* 115: 203-213.

NÖTH, WINFRIED. 2000a. *Handbuch der Semiotik*, 2. Aufl. Metzler.

__. 2000b. Selbstreferenz in systemtheoretischer und in semiotischer Sicht. Em: A. Barsch, G. Rusch, R. Viehoff e F. W. Block, (eds.) Festsite Siegfried J. Schmidt. http://www.sjschmidt.net/konzepte/texte/noeth1.htm - Trad.: Auto-referência na teoria dos sistemas e na semiótica. *Revista de Comunicação e Linguagens* [Lisboa] 29 (2001): 13-28.

__. 2001a. Protosemiotics and physicosemiotics. Em: The Semiotics of Nature. *Sign Systems Studies* 29 (1): 13-26.

__. 2001b. Semiosis and the Umwelt of a robot. *Semiotica* 134: 695-699.

__. 2002. Semiogenesis in the evolution from nature to culture. Em: *Origin of Semiosis*, Patrizia Violi (ed.). Brepols.

PAPE, HELMUT. 1993. Final causality in Peirce's semiotics. *Transactions of the Charles S. Peirce Society* 29: 581-607.

PATTEE, HOWARD H. 1997. The physics of symbols and the evolution of semiotic controls. Em: *Control Mechanisms for Complex Systems*.

Michael Coombs e Mark Sulcoski (eds.) Univ. of New Mexico Press, 9-25.(Also:http://www.ssie.binghamton.edu/pattee/semiotic.html)

PEIRCE, CHARLES SANDERS. 1887. Logical machines. *American Journal of Psychology* 1.1: 165-170. [Também em: C. S. Peirce, 2000, Writings, vol. 6, ed. N. Houser et al., Bloomimgton: Indiana Univ. Press, 65-72.]

__. 1931-1958. *Collected Papers*. Vols. 1-6, eds. C. Hartshorne & P. Weiss, vols. 7-8, ed. A. W. Burks. Harvard Univ. Press.

POPPER, KARL. 1972. *Objective Knowledge*. Oxford Univ. Press.

SANTAELLA, LUCIA. 1994. Peirce's broad concept of mind. S: *European Journal for Semiotic Studies* 6: 399-411.

SANTAELLA, LUCIA. 1998. Der Computer als semiotisches Medium. Em: *Medientheorie und die digitalen Medien*. Winfried Nöth e Karin Wenz (eds.). Kassel Univ. Press, 121-158.

__. 1999. A new causality for the understanding of the living. *Semiotica* 127: 497-518.

SCHMIDT, SIEGFRIED J. 1987. Der radikale Konstruktivismus. Em: *Der Diskurs des radikalen Konstruktivismus*. S. J. Schmidt, ed. Suhrkamp, 11-88.

SEARLE, JOHN. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3: 417-457.

SKAGESTAD, PETER. 1993. Thinking with machines: Intelligence augmentation, evolutionary epistemology, and semiotic. *Journal of Social and Evolutionary Systems* 16.2: 157-180.

__. 1999. Peirce's inkstand as an external embodiment of mind. *Transactions of the Charles S. Peirce Society* 35: 551-561.

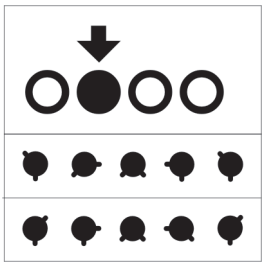
TIERCELIN, CLAUDINE. 1993. *La pensée signe:* Études sur C.S. Peirce. J. Chambon.

WINOGRAD, TERRY; FLORES, FERNANDO. 1886. *Understanding Computers and Cognition*. Norwood, N.J.

COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA



CAPÍTULO 7

UM ROBÔ POSSUI *UMWELT*? REFLEXÕES SOBRE A BIOSSEMIÓTICA QUALITATIVA DE JAKOB VON UEXKÜLL

Claus Emmeche

INTRODUÇÃO

Como o conceito de *Umwelt* de Jakob von Uexküll dialoga com as discussões mais recentes de biologia teórica, filosofia da biologia, biosemiótica e vida artificial, particularmente as pesquisas sobre 'sistemas autônomos' e robôs? Para investigar esta questão, minha abordagem não se desenvolverá como uma retrospectiva histórica das idéias de Uexküll, expondo o núcleo original de idéias filosóficas que serviram de infra-estrutura ao conceito original de *Umwelt* (alguns parecem incompatíveis com uma perspectiva evolucionista moderna). Ao contrário, mostrarei que alguns aspectos de suas propostas são ainda interessantes e podem trazer inspirações à biologia contemporânea, à ciência cognitiva e a outros campos. Além disso, chamarei a atenção para as reflexões de Thure von Uexküll em seu desenvolvimento da teoria do *Umwelt*, que é notadamente evolucionário (como a abordagem de seu pai).

Mais especificamente, vou investigar a plausibilidade de três teses: (I) a teoria do *Umwelt*, de Jakob von Uexküll, pode, no contexto da ciência contemporânea, ser interpretada mais adequadamente como um ramo do *organicismo qualitativo* dentro da biologia teórica, mesmo que sua biologia teórica tenha sido freqüentemente caracterizada como sendo eminentemente vitalista. O organicismo qualitativo é 'uma posição em cima do muro', ou seja, por um lado afirma que não há poderes vitais misteriosos, ou não-materiais, em organismos (não vitalismo), mas por outro lado afirma que as propriedades características dos seres vivos não podem ser totalmente explicadas nem pela física e nem pela química, uma vez que estas propriedades são não-redutíveis e emergentes (emergentismo). Segundo, que algumas destas propriedades emergentes tem um caráter experiencial, fenomenal ou subjetivo que desempenha um papel fundamental na dinâmica do sistema vivo. A biossemiótica moderna (inspirada por C.S.Peirce e J. von Uexküll) é um tipo de organicismo qualitativo. (II) Esta posição ilumina algumas discussões recentes em ciência cognitiva, vida artificial e robótica sobre a natureza da representação e da cognição. De fato, há questões genuinamente semióticas, uma vez que lidam com o papel da informação e dos signos, em qualquer sistema que tenha a propriedade de ser 'como um animal', ou seja, sistemas que se movem e parecem guiados por algum tipo de intelecto ou, utilizando outra terminologia, um programa comportamental. (III) Particularmente, o organicismo qualitativo nos permite abordar a questão sobre se um robô *pode* ter um *Umwelt*, no sentido exato que Jakob von Uexküll atribuía a este termo (um mundo fenomenal subjetivamente experienciado). Uma resposta positiva a esta questão, ou seja, a afirmação de que um robô realmente pode ter um *Umwelt*, parece ser contra-intuitiva, uma vez que um robô pode ser visto como uma *encarnação* da visão mecânica e reducionista de mundo a qual Jakob von Uexküll se opunha fortemente. Mas certas idéias e conceitos podem nos levar a conseqüências inesperadas, que confrontam nossas suposições metafísicas, de forma que deveríamos tentar encarar tais questões de mente aberta.

Assim, ao elaborar esta terceira questão, devemos também saber se isso é o mesmo que perguntar: Um robô pode ter uma mente? Se sim, o conceito de *Umwelt* é apenas outro conceito para mente, e a teoria de Jakob von Uexküll não nos ajudaria a resolver nossa questão. Mas este não é o caso. Apesar disso, poderíamos pensar que, se utilizarmos um conceito muito amplo de mente, por

exemplo, motivado pela biossemiótica e pela filosofia de Peirce, então o *Umwelt* e a mente de animais devem ter a mesma extensão. Entretanto, uma mente, tal como se encontra em Peirce (Santaella Braga 1994), não é co-extensiva de *um Umwelt*. A noção de mente é mais ampla do que a idéia de *Umwelt*. Assim, podem haver muitas atividades em um organismo vivo que são de caráter mental (ou semiótico), mas não aparecem como parte do mundo fenomenal do animal. Os dois conceitos significam coisas diferentes. Eu não saberia dizer se Peirce teria associado propriedades mentais a robôs, mas parece-me ser o caso que o faria — tanto organismos biológicos como robôs com sensores e atuadores poderiam, em princípio, incorporar os mesmos princípios lógicos ou semióticos (cf. Burks 1975).

Este capítulo segue o seguinte percurso. Após uma breve introdução ao conceito de *Umwelt*, de Jakob von Uexküll, sua teoria será situada na tradição do organicismo qualitativo na biologia. Permitam-me enfatizar que a teoria do *Umwelt* pode ser interpretada de outras maneiras (por exemplo, como sendo estritamente vitalista), de forma que o que pretendo fazer aqui não é uma exposição crítica da versão de Uexküll mas uma reconstrução de sua teoria mais alinhada com a biologia teórica contemporânea. O próximo passo é uma visão histórica da pesquisa em robótica e sistemas autônomos, um campo científico que já atraiu a atenção dos semioticistas (cf. Meystel 1995) e que é profundamente inspirado em considerações biológicas. Mais adiante, a questão sobre se um robô pode ter um *Umwelt* será finalmente decidida (ou pelo menos assim se espera). As perspectivas sobre uma maior aproximação entre biologia teórica, semiótica, pesquisa de sistemas autônomos e ciência cognitiva serão discutidas.

O CONCEITO DE *UMWELT* E A BIOLOGIA NOS DIAS DE HOJE

Umwelt: não é ambiente, não é mente

Umwelt pode ser definido como o aspecto fenomenal das partes do ambiente de um sujeito (um organismo animal); as partes que ele escolhe com órgãos sensoriais específicos de sua espécie, de acordo com sua organização e suas necessidades biológicas (J. von

Uexküll 1940, T. von Uexküll 1982a, 1989). Neste sentido, o sujeito é o construtor de seu próprio *Umwelt*, uma vez que tudo nele está catalogado a partir de sinais perceptivos e de atuação do próprio sujeito. Deve-se pelo menos distinguir entre os seguintes conceitos: (1) o *habitat* do organismo como 'objetivamente' (ou externamente) descrito por um observador científico humano; (2) o nicho do organismo, no sentido ecológico tradicional, como a função ecológica das espécies dentro de um ecossistema, (3) o *Umwelt*, como o mundo experienciado pelo organismo.[1]

A noção de *Umwelt* influenciou profundamente Konrad Lorenz mas nunca se estabeleceu realmente e foi esquecida por um longo período. Isto pode ser, em parte, devido à predominância do pensamento Darwiniano na biologia, e ao fato de que Jakob von Uexküll tinha, desde muito cedo, se tornado um convicto anti-Darwinista e era também conseqüentemente associado à oposição entre vitalismo e mecanicismo na biologia (ver Harrington 1996. Lorenz, é claro, era um Darwinista. Como Richards (1987: 530) observa:

> Apesar do compromisso adamantino de Lorenz com o ultra-Darwinismo, sua teoria do instinto foi construída sobre as idéias de um pensador abertamente anti-Darwinista - Jakob von Uexküll, um vitalista Drieschiano. De von Uexküll, um acadêmico independente de dedicação integral, Lorenz adaptou a noção de 'um sistema funcional' (*Functionskreis*). De acordo com a teoria de von Uexküll, um sistema funcional ou interativo é constituído pela relação entre um animal, seus órgãos e necessidades especiais, e seu mundo experienciado (*die Umwelt*), cuja realidade vivida corresponderia às habilidades sensoriais e às exigências do animal. Lorenz transformou o conceito de von Uexküll de sistema funcional em um 'esquema de emissão inata' (*angeborenen Auslîse-Schemata*). Este mecanismo de emissão inata, como ele também o denominou, seria o correlato receptivo no animal que responde com um padrão particular de comportamento a sinais específicos no ambiente.

Esta passagem ilumina as razões históricas pelas quais o foco no aspecto fenomenal foi rapidamente perdendo sua importância e quase desapareceu nos desenvolvimentos subseqüentes relacionados à vertente principal do estudo de comportamento animal, provavelmente por causa das influências de movimentos como o positivismo, behaviorismo, e, na biologia, o neo-Darwinismo e o mecanicismo.

Entretanto, com o desenvolvimento na segunda metade deste século da zoosemiótica, da biosemiótica e da medicina psicossomática, a noção de *Umwelt* voltou a ser usada. Basta esta

citação de um artigo sobre 'Endosemiose', de Thure von Uexküll, Wernes Geigges e Jîrg M. Hermann, para reafirmar a base deste conceito: 'Jakob von Uexküll cunhou o termo *Umwelt* ("universo subjetivo", "entorno significativo", "mundo fenomenal", ou "automundo", como oposto ao termo "ambiente", *Umgebung*' (T. von Uexküll et al. 1993: 6). Em uma nota a esta citação, os autores propõem que *Umwelt* é o mundo subjetivo com influência significativa sobre um ser vivo, considerando seu aparato de processamento de informação, seu sistema de signos, e seus códigos. Eles continuam, observando que os animais estão envoltos em redes de processos sígnicos que os protegem, traduzindo o ambiente em seu significado subjetivo, acessível apenas ao sujeito que codifica.

Duas características da noção de *Umwelt* são importantes neste contexto. (i) Como observado, o *Umwelt* de um sistema não pode ser identificado com uma mente. Qualquer que seja a forma pela qual se caracterize uma mente, sua atividade é mais abrangente do que aquilo que é experimentado pelo sistema como sendo seu mundo. Por exemplo, em humanos o *Umwelt* se torna consciente por meio da percepção intencional, da cognição e da linguagem, enquanto um oceano de processos subconscientes ou inconscientes são também partes ativas da mente. (ii) Um organismo tem somente um acesso primário ao seu próprio *Umwelt*, e apenas humanos (e alguns animais leitores de mentes tais como predadores capazes de interpretar a mente de suas presas) podem, por inferência, ter acesso indireto ao *Umwelt* de outras espécies. Entretanto, esse 'acesso indireto' nunca é exatamente o *Umwelt* real das espécies. Por exemplo, nossa compreensão científica do sistema de sonar de um morcego nos dá uma idéia indireta e funcional do *Umwelt* do morcego, mas não podemos 'entrar' no *Umwelt*; tudo o que temos é um modelo, em nosso *Umwelt* (lingüístico, cognitivo e perceptivo), do *Umwelt* do morcego. A ciência procura construir uma 'visão não localizada' (Nagel 1986), baseada em modelos, mas só é capaz de fazer isto mediada pelo *Umwelt* específico de nossa espécie, nosso ponto de vista a partir do qual coletivamente construímos uma esfera de conhecimento público compartilhável.

Organicismo qualitativo

Uma interpretação equivocada, que pode ser freqüentemente encontrada, refere-se a construção da filosofia da biologia do século XX como uma luta entre o vitalismo e o mecanicismo, que teria

sido finalmente ganha pelo mecanicismo. Esta elaboração ignora o fato de que a posição mais influente passou a ser organicista (embora a ciência popular, depois dos triunfos da biologia molecular, tenha contado uma história diferente para o público). A 'solução para o debate' entre vitalismo e mecanicismo não foi uma posição mecanicista, mas um tipo de compromisso histórico na forma do que chamo de *corrente principal do organicismo* (exemplificado pelos textos de biólogos bem conhecidos como J. Needham, P.Weiss, C.H. Waddington, J. Woodger, E. Mayr, R.C. Lewontin, R. Levins, S. J. Gould) funcionando mais ou menos tacitamente como uma base filosófica da biologia. O organicismo toma a complexidade e a singularidade do organismo como signo da distinção da biologia como ciência natural.[2] Essa posição tem várias raízes históricas. Um precursor é o movimento emergentista do início do século XX, especialmente na Grã- Bretanha.[3] Este impasse, embora aqui tratado sob uma perspectiva naturalista evolucionária, foi antecipado pela noção de Kant, mais crítica (não naturalista), de organismo vivo.[4] Segundo Kant, não podemos dispensar um princípio heurístico de propósito quando consideramos um organismo — 'Um produto organizado da natureza é aquele no qual cada parte é reciprocamente um propósito [fim] e um meio. Nele, nada é em vão, sem propósito, ou determinado por um mecanismo cego da natureza' (Kant 1790 [1951: 222]). Entretanto, dentro da principal corrente do organicismo, esta teleologia é interpretada mais ou menos como uma teleonomia 'mecânica', o resultado de forças de variação cega e seleção natural, talvez eventualmente uma 'ordem livre' adicional, ou auto-organização física. Desta forma, a corrente principal organicista é não vitalista, ontologicamente não reducionista (permitindo redução metodológica) e emergentista. O que se estuda como propriedades emergentes são estruturas materiais comuns e processos dentro de vários níveis de sistemas vivos (sistemas de desenvolvimento, evolução, propriedades de auto-organização, etc.); todos são tratados como objetos sem nenhuma propriedade experiencial intrínseca. Por exemplo, em estudos comportamentais, os etologistas não têm permissão para usar uma linguagem subjetivista, para descrever um comportamento animal.

Em contraste, o *organicismo qualitativo* representa uma visão mais 'colorida' dos seres vivos. Ele enfatiza não apenas a realidade ontológica de propriedades ou entidades biológicas de nível mais alto (tais como o sistema de organismos auto-reprodutores) mas também a existência de aspectos fenomenológicos, ou qualitati-

vos, de pelo menos algumas propriedades de nível mais alto. Quando está sentindo a luz, ou as cores, um organismo não está meramente detectando sinais externos que são processados internamente, que poderia ser descrito em termos neuroquímicos, de 'processamento de informação', ou algo assim. Se queremos a história completa, algo mais está acontecendo. Embora a experiência tenha um modo subjetivo de existência, ela é um fenômeno objetivamente real (na filosofia da mente mais recente, Searle [1992] é um dos poucos a enfatizar a realidade ontológica da experiência subjetiva. Entretanto, ele está na maior parte do tempo falando sobre a experiência humana). Como posição científica, o organicismo qualitativo se preocupa não apenas com qualidades, como as categorias das qualidades 'primárias' (que corresponderia ao *quanta* do que é cientificamente mensurável) incluindo forma, magnitude, e número, mas também com as qualidades secundárias de cor, sabor, tato, som, sentimento etc.[5] Não se deve tentar equiparar o organicismo qualitativo, ou a corrente principal do organicismo, a posições, teorias ou paradigmas coerentes, embora possam ser encontrados, em ambos, representantes na biologia teórica recente.[6] Alguns autores podem não ser consistentes, alguns podem apenas implicitamente expressar sua idéia; o importante é reconhecer que dois conceitos diferentes de vida e biossemiose estão em jogo.

É obvio que a noção de *Umwelt* é de importância central para o desenvolvimento de uma teoria coerente do mundo qualitativo e experiencial do organismo, uma tarefa que a biologia atual precisa enfrentar, ao invés de continuar ignorando um enorme escopo fenomenal do mundo vivo — o mundo experiencial dos apetites animais, dos desejos, dos sentimentos, das sensações, etc.[7] Para tal tarefa, pode-se encontrar inspiração teórica nos campos da semiótica, bem como em vida artificial e em pesquisas sobre sistemas autônomos. O *Umwelt* está enraizado no corpo material e semiótico do organismo, que está situado em um nicho específico. A teoria não deve postular quaisquer poderes ocultos, vitalistas, ou espirituais ao 'explicar' a emergência dos *Umwelt*s na evolução; entretanto, deve conhecer a riqueza e a realidade do fenômeno de sentir, agir e perceber do organismo. A implicação de tal aventura poderia ser importante não apenas para a biologia, mas também para a semiótica (para estabelecer a noção de signo na natureza), para a filosofia da mente (para superar o dualismo e resolver os problemas do fisicalismo superveniente não redutível), e para o

entendimento geral da relação entre o ser humano e as outras espécies. Poderíamos criar um 'organismo' artificial com um *Umwelt* mais estranho para nós do que aquele de um chimpanzé ou da mosca das frutas?

SISTEMAS AUTÔNOMOS: UMA BREVE HISTÓRIA

Freqüentemente se sugere que dispositivos voltados para propósitos humanos específicos poderiam ser mais convenientes e úteis se pudessem ser 'agentes autônomos'. Isto é, se não fossem apenas dispositivos computacionais de entrada e saída, mas fossem comandados por sistemas cibernéticos com seus próprios módulos motores guiados por sensores, tomando decisões, tendo a capacidade de agir de forma mais ou menos inteligente, dada uma informação parcial, aprendendo com erros, adaptando-se a ambientes mutáveis e heterogêneos, e tendo, assim, um tipo de vida própria. Sem dúvida, tais dispositivos também poderiam causar prejuízos, mas irei desconsiderar preocupações quanto ao acesso à tecnologia e suas implicações éticas. Vários programas de pesquisa têm sido propostos e estão voltados para o estudo do que se costuma chamar de 'sistemas autônomos', 'agentes situados', 'sistema IA distribuídos' e 'sistemas multi-agentes'.

Este campo de pesquisa, denominado Pesquisa em Sistemas Autônomos (PSA), é um desenvolvimento da área clássica de Inteligência Artificial (IA) em vários aspectos, especialmente em seu *estruturalismo implícito*: o objetivo não é tanto o estudo científico de formas naturais do fenômeno (comportamento inteligente), mas sim sua estrutura processual mais abstrata e geral, obviamente para ver se outras instâncias de sua estrutura poderiam ser desenvolvidas artificialmente de forma a resolver alguns problemas específicos. No caso da IA clássica, das décadas de 1950 e 1960, o objetivo não era o estudo científico da inteligência *humana*, que veio a se tornar o foco da ciência cognitiva (CC), mas a criação de um conjunto de teorias de *sistemas inteligentes possíveis*, que pudessem ser implementados em instâncias físicas de máquinas de Turing. Com o advento da Vida Artificial (VA), como programa de pesquisa, ao final da década de 80, o propósito teórico passou a ser o estudo da 'vida como ela poderia ser', de forma a se estender, por assim dizer, o conjunto-base de exemplos fornecidos pela biologia 'carbono-chauvinista' tradicional, acusada de lidar empiricamente

com apenas uma única classe de sistemas vivos — aqueles que acidentalmente evoluíram na Terra.[8] Por outro lado, o estudo de sistemas autônomos não se concentra na estrutura causal de sistemas autônomos naturais (micróbios, plantas, animais e seres humanos) mas na estrutura de qualquer sistema concebível que possa realizar comportamento autônomo.

Este estruturalismo pode ser visto como algo importante para o projeto criativo e para a engenharia de novos tipos de sistemas, uma necessária libertação do foco na investigação empírica de sistemas realizados na natureza. Entretanto, na IA, e certamente na PSA, ele criou uma confusão epistemológica em relação à *síndrome de pigmaleão* (cf. Emmeche 1994a:63, 134-155). Esta é a falácia que surge quando se assume que um modelo criado artificialmente não somente *representa* a realidade mas *se torna efetivamente* uma instância alternativa da realidade. Se um sistema de vida artificial, tal como o TIERRA de Tom Ray (ver Langton et al. 1992), é tomado não só para modelar alguns aspectos abstratos da evolução por seleção natural, mas para ser uma instância de vida, comete-se a falácia de pigmaleão. Há um extenso debate sobre o que é realmente uma falácia em um programa de pesquisa em VA 'forte'.[9] De maneira análoga, se poderia postular que os dispositivos criados dentro do campo da PSA são *modelos* mais ou menos interessantes de organismos autônomos vivos 'reais' (onde sistemas artificiais não são intrinsecamente autônomos, uma vez que a propriedade de autonomia é atribuída a eles na medida em que funcionam como modelos), ou que são simplesmente máquinas cibernéticas que certamente poderiam se comportar *como se fossem* autônomos, mas cuja autonomia é, ou simples demais para capturar a propriedade desejada da coisa real, ou simplesmente de uma outra categoria de comportamento. O conceito de *Umwelt* raramente aparece nessa discussão[10], embora uma real compreensão do conceito de *Umwelt* pudesse ter profundas implicações para a VA forte.

Meu ponto aqui não é dizer que as pesquisas em sistemas autônomos, *animats* e robótica fracassaram porque tais sistemas nunca serão 'verdadeiramente autônomos', ou porque são biologicamente irrealísticos, inúteis, excêntricos, ou similares — isso pode ser fato para alguns ou para todos os sistemas criados até hoje, mas deve-se reconhecer um rico e variado campo de pesquisa e desenvolvimento tecnológico, que pode inspirar não só a nascente indústria de 'projetos inteligentes' mas também um novo leque de investigações científicas. É ainda umaquestão em aberto o que poderá ser

obtido por essas pesquisas no futuro. Meu ponto aqui é articular algumas questões, ao invés de tentar obter respostas. Uma questão crucial é se tais sistemas podem ou não ter um *Umwelt*, e se podem, como este seria. Além disso, como saberíamos disso, e se não pudéssemos saber, por que não poderíamos? Uma abordagem destas questões envolve investigação de pressupostos metafísicos. Que tipos de sistema são realmente autônomos? São aqueles com uma relação intrínseca entre *Umwelt* e autonomia?

Uma observação sobre terminologia. O termo 'autônomo', é empregado de muitas maneiras na literatura em PSA, muito freqüentemente com um significado informal. Autonomia significa, no senso comum, uma pessoa, uma região ou um estado que é capaz de se auto-governar, é independente e sujeito a suas próprias leis.[11] Um significado possível é 'liberdade', tal como a liberdade da vontade (livre-arbítrio). Tais significados dependem de que tipo de sistema é visto como sendo autônomo.[12] O termo *autônomo* deriva da palavra grega auto-, ou *autos*, significando o próprio, a si-mesmo, e nomos, significando lei; ou seja, auto-governado, auto-dirigido, espontâneo, oposto a *heterônomo*, i.e., que é controlado externamente. Na teoria biológica de Maturana e Varela (1980), o termo ganhou um significado especial — a condição de subordinar todas as mudanças à manutenção da organização e 'a capacidade de auto-manutenção de sistemas vivos, no sentido de manterem sua identidade por meio de compensações ativas a deformações' (ibid.: 135).[13] Entretanto, em PSA, em um grande número de casos, aquilo que se entende por 'agente autônomo' é classificado como um sistema não-autônomo (heteropoiético) de acordo com os critérios da teoria de autopoiese.

Cibernética, robótica, IA clássica: alguns antecedentes históricos

A noção de sistemas autônomos tem origem tanto em idéias pré-científicas sobre comportamento adaptativo, inteligente, orientado para obter soluções, em homens, animais e máquinas, quanto em tentativas iniciais para modelar e construir sistemas com um comportamento que parece orientado para metas. Elas foram desenvolvidas durante o período inicial da cibernética, da teoria da informação e de disciplinas correlatas (teoria de sistemas, teoria da operação). Outros campos também foram cruciais, posteriormente: raciocínio automático, reconhecimento de padrões, bancos

de dados inteligentes, sistemas especialistas, e a robótica, no estilo clássico da IA. A história da origem e do intercâmbio de idéias entre as diferentes disciplinas do movimento do 'pensamento sistêmico' (*systems thinking*) é bastante relevante para que se possa compreender o *background* histórico da PSA, embora complicada demais para tratarmos aqui.[14] Entretanto, a *cibernética* merece a nossa atenção, em parte porque a PSA pode ser vista como uma extensão de alguns aspectos do programa original de pesquisas em cibernética, e, parcialmente, porque alguns aspectos da cibernética podem ser vistos como uma versão mecanicista do ciclo funcional da teoria do *Umwelt*.

A idéia de uma arte, e de uma ciência do controle, entre uma vasta gama de campos onde esta noção pode ser aplicável, foi proposta pelo matemático Norbert Wiener, em 1948. A cibernética é uma teoria de sistemas com realimentação (retro-alimentação ou *feedback*), ou seja, sistemas auto-regulados tais como máquinas e animais. A noção central é a de realimentação, ou seja, a alimentação recursiva de informações sobre alguma mudança em parâmetros descrevendo o estado de uma parte de um sistema (por exemplo, alguma medida de saída ou de desempenho) para os mecanismos responsáveis por efetuar essas mudanças, freqüentemente com a função de regular o comportamento do sistema a fim de mantê-lo em uma região confortável de interação com o meio ambiente (realimentação negativa).[15]

Um exemplo significativo é o sistema sensor-perceptivo-(cognitivo)-motor de nosso corpo. Quando nos movemos para pegar uma bola, interpretamos nossa visão do movimento da bola para prever sua trajetória. Nossa tentativa de alcançar a bola envolve a antecipação de seu movimento de modo a determinar o movimento de nosso corpo. Quando a bola se aproxima, percebemos que ela está afastada da trajetória esperada e ajustamos nosso movimento. Na descrição cibernética, o sujeito é descrito como um mecanismo de processamento de informações. Assim o sistema visual pode ser visto como algo que provê entradas para um controlador (e.g. cérebro), que por sua vez deve gerar sinais de controle para fazer com que o sistema motor (e.g. músculos) se comporte da forma desejada (e.g. pegar uma bola). A alimentação-direta (*feedforward*) antecipa a relação entre o sistema e o ambiente para determinar um curso de ação; a realimentação (*feedback*) monitora as discrepâncias que podem ser usadas para ajustar as ações. Assim, o

problema do controle é escolher a entrada adequada de forma que o sistema possa determinar uma saída correspondente, de modo que se comporte de uma maneira desejada, ou mantendo a saída próxima a um valor de referência (problema da regulação), ou fazendo com que a saída siga uma trajetória desejada (problema de servo-controle). Enquanto um sinal de controle, definido por seu efeito de antecipação, não conseguir atingir este efeito, uma realimentação é necessária para comparar esta antecipação com o atual estado, e desta forma determinar uma mudança que compense esta diferença. Uma sobre-compensação acaba levando o sistema à instabilidade; uma subcompensação leva a um ajuste pobre e a um desempenho lento, com atrasos. Desta forma, os princípios da cibernética são relativamente fáceis de descrever em alto nível, uma vez que podemos compreender intuitivamente os mecanismos usados para alcançar uma bola, por exemplo, jogando tênis. Matematicamente estes princípios são mais difíceis de analisar. Simular este tipo de comportamento, em alto nível de complexidade, e em tempo real, é computacionalmente difícil. E ninguém foi ainda capaz de projetar um agente autônomo que possa imitar, nem mesmo com uma 'pequena parcela de graça', o ato natural de uma pessoa que intercepta uma bola, por exemplo, quando joga tênis.

A descrição cibernética da informação sendo realimentada repetidas vezes entre os componentes do sistema tem foco no papel dos signos individuais, que von Uexkull chamou de 'ciclo funcional'. A linguagem mais teórica do *Umwelt* pode ser traduzida para a linguagem da cibernética sem perda de significado. Entretanto, a cibernética corresponde a uma descrição externalista. Ela desconsidera que um mundo subjetivo possa ser experienciado por um organismo a partir de um ponto de vista interno. Dessa forma, a 'informação' da realimentação cibernética não deve ser associada ao conceito de signos perceptivos e operacionais do ciclo funcional. Estes últimos podem ser mais adequadamente interpretados como conceitos semióticos envolvendo relações triádicas entre signos, objetos e interpretantes.[16] Neste sentido, uma tradução que preserve o significado de maneira completa pode não ser possível, e estes dois modos de descrição são parcialmente incompatíveis. Essa distinção é importante porque qualquer dispositivo simples, que significativamente possa ser descrito como processando sinais, é controlado do ponto de vista informacional, segundo este sentido

mais simples, e mesmo assim tal dispositivo pode não possuir um *Umwelt*.

O que aconteceu com a cibernética? Hoje, após a introdução dos computadores, os estudos teóricos de problemas de controle têm se tornado tão sofisticados, e suas aplicações (na engenharia, na biomedicina, na economia e, claro, na robótica e na IA) têm se enraizado de maneira tão firme, que é difícil resgatar o furor intelectual decorrente das idéias de Wiener. Além disso, depois da revolução cognitiva na década de 60, uma ênfase maior foi atribuida às capacidades cognitivas de alto nível, que eram intratáveis, não-gerenciáveis por meio de princípios puramente cibernéticos, uma vez que pareciam pressupor a ação de sistemas simbólicos extensivos para raciocinar e representar a informação a respeito da natureza das tarefas a serem resolvidas. Isso iniciou o desenvolvimento da Ciência Cognitiva (CC), e antes disso, da IA. Embora seja precipitado considerar um esquema histórico válido, não se pode descartar que o interesse por sistemas autônomos fora representado, no início, pela cibernética e pela ciência dos sistemas; depois, por volta de 1950 e 1960, pelo novo campo da IA e da robótica; nos últimos 30 anos, pelas transformadoras regiões de encontro entre IA, CC, robótica, neurociência, e recentemente Vida Artificial, biologia teórica, além da PSA. Isso não significa que a cibernética esteja 'morta' ou que pesquisas não existam sob o rótulo da cibernética (e.g. Heylighen *el al*. eds. 1990); significa que todo o campo mudou muito com os desenvolvimentos de pesquisas em sistemas complexos, CC, IA, etc. Princípios cibernéticos estão fortemente integrados no núcleo das pesquisas em PSA. Vamos examinar a noção de autonomia, de um ponto de vista mais próximo da robótica tradicional, para entendermos contra o que novos estudos sobre 'cognição incorporada' e o movimento PSA estão se colocando.

A boa e velha robótica

Apesar da existência de linhas distintas nos programas atuais de pesquisa em robótica — do desenvolvimento de melhores sistemas de controle motor, para realização de tarefas simples, pré-definidas e úteis em linhas industriais de montagem, às mais ambiciosas, como corporificar sistemas de inteligência geral para servirem seres humanos — a pressuposição geral da robótica, disseminada nos anos 80, é que a atividade baseada em conhecimento pode ser inteligente apesar de mediada por uma máquina. Na prática, siste-

mas de IA podem ser autônomos, como robôs, ou podem ser amplificadores de inteligência, quando usados para aumentar o desempenho humano na tomada de decisões.[17] Sistemas robóticos de IA devem ser capazes de reconhecer objetos ou cenas (como um serviçal real[18] pode fazer) e interagir com o mundo, através de câmeras ou simples sensores de toque.

Tal sistema robótico deve ter capacidade de aprendizagem. Ele deve aprender pela extração de características visuais úteis de informações que recebe e calibrar seu 'espaço visual' representado internamente através da exploração pelo toque ou visão de objetos no mundo. Tais dispositivos robóticos[19] empregam reconhecimento de padrões juntamente com o conhecimento armazenado (freqüentemente representado de forma estável e simbólica) a fim de inferir (a partir de sinais de entrada e do conhecimento armazenado do seu mundo de objetos) as formas tridimensionais e propriedades não percebidas dos objetos, mesmo que os dados sejam limitados e não estritamente adequados. Pelo menos este era, e de certa maneira continua a ser, o objetivo ambicioso de construção de tais sistemas. Sua percepção de padrões não é, normalmente, flexível como deveria ser, sendo a ênfase no uso de dados sensoriados em tempo real por programas que usam conhecimento armazenado. Isto apresenta falhas em situações atípicas, já que as inferências sobre o mundo dependem de pressuposições apropriadas, e o que é 'apropriado' é altamente dependente do contexto e da situação como um todo, e a base de conhecimento do robô é restrita a poucas situações de micro-mundos. Isto é uma parte do *frame problem* em IA; não um detalhe técnico mas um sério obstáculo ao projeto de qualquer tipo de sistema que modela um mundo complexo e em mudança (para detalhes, veja Janlert 1987).

A atual pesquisa em IA é mais orientada para o desenvolvimento de programas que sofisticados *hardwares*. Os dispositivos robóticos podem servir de laboratórios para sugerir e testar programas. No entanto, a IA não é somente engenharia de comunicação avançada ou programação em lógica. Além do objetivo de construir ferramentas 'inteligentes', a IA pode ser (mas não precisa ser) vista como uma declaração sobre a natureza da mente. O que John Haugeland (1985) batizou de GOFAI (*Good Old Fashioned Artificial Intelligence*) é a afirmação de que (a) nossa habilidade para lidar com as coisas de forma inteligente é devido a nossa capacidade de pensar sobre elas de maneira racional (incluindo o pensamento subconsciente); (b) nossa

capacidade de pensar sobre as coisas de maneira racional equivale à faculdade de manipulação interna 'automática' de símbolos, atuando sobre um conjunto de representações estáveis armazenadas. Isto implica que as manipulações internas de símbolos devem ser interpretadas como sendo sobre o mundo exterior, e que as manipulações 'racionais' internas de símbolos devem ser conduzidas por algum subsistema computacional ('computadores internos'). Isto não é somente de interesse filosófico, já que este paradigma, quando aplicado à arte de construir robôs, cria uma figura de robô como um veículo com um computador avançado de IA embarcado, por exemplo um imenso sistema especialista (onde a especialidade idealmente deve ser de senso comum!) equipado com sensores e atuadores.

Este estilo de IA, tradicional de pesquisa com foco no conhecimento explícito, escolhas racionais e solução de problemas, tem provado ser de difícil aplicação na construção de robôs autônomos. Os poucos sistemas construídos mostram deficiências como fragilidade, inflexibilidade, pouca operatividade em tempo real, etc. Os problemas que apareceram em IA, neste contexto, como o problema de raciocínio não monotônico, e o *frame problem* (*Pylyshyn* 1987), são de claro interesse teórico, mas permanecem sem solução. E as soluções sugeridas não parecem ser particularmente úteis para o desenvolvimento de sistemas situados. Outra característica da robótica é a tradicional abordagem de cima-para-baixo *(top-down*). Nenhum dos módulos, considerados isoladamente, gera o comportamento total do robô. É preciso combinar vários módulos para obter um comportamento qualquer do sistema. Melhorias na performance do robô vêm da melhoria dos módulos funcionais individualmente. Isto é difícil, em razão da inflexibilidade das competências funcionais das várias partes, uma vez que as mudanças em um módulo irão afetar negativamente a performance de outro, de forma que o projeto completo precisa ser reconsiderado, a cada passo da mudança (este problema é, de certa forma, remediado pelas novas abordagens em agentes autônomos). As ênfases na explicitação do conhecimento, do raciocínio e do projeto de cima-para-baixo são irrealistas do ponto da vista da biologia, do comportamento e *Umwelt* do animal real.

A Boa e Velha Robótica herda a *Hipótese do Sistema Físico de Símbolos* (*PSSH — Physical Symbols Systems Hypothesis*) da IA. Esta hipótese[20], que está muito longe da biologia real, afirma que os processos necessários para produzir comportamento inteligente

podem ser simulados com uma coleção de símbolos físicos e um conjunto de mecanismos que produzem séries, ao longo do tempo, de estruturas construídas a partir destes símbolos. O computador digital deve funcionar como uma ferramenta com a qual as estruturas simbólicas são formadas e manipuladas. Estruturas simbólicas em um programa de IA são usadas para conhecimento geral sobre o domínio de um problema (como jogar xadrez, realizar diagnóstico médico, ou, mais relevante para robôs autônomos, realizar distinções entre objetos, criar categorias etc.) e para especificar o conhecimento sobre a solução do problema corrente. Por que sistemas de símbolos devem ter um papel necessário na ação inteligente? Da perspectiva da robótica em IA (cf. Newell 1980)[21], (a) a racionalidade demanda designação de situações potenciais, (b) sistemas de símbolos provêem isso, (c) somente sistemas de símbolos podem prover isso quanto novidade e diversidade suficientes de tarefas são permitidas.

Portanto, a idéia implícita é que as interfaces perceptivo-motoras são conjuntos de símbolos. O sistema central opera de maneira independente deste domínio (símbolos). Seu significado não é importante para aquele que efetua o raciocínio, mas a coerência do processo como um todo emerge quando (1) o observador do sistema conhece a fundamentação dos símbolos dentro de sua própria experiência[22], ou (2) o sistema funciona tão bem, como um todo (engenho de raciocínio e os módulos sensório-motor), que constitui o local de *emergência de significado* no sentido de funcionamento bem adaptado. Implicada, no paradigma simbólico, está a idéia de que os símbolos, e suas concatenações, representam entidades no mundo, sejam elas coisas individuais, propriedades, conceitos, estados intencionais de outros agentes, qualidades perceptivas, etc. A inteligência central do robô 'processa' símbolos alimentados pelo sistema perceptivo. É fornecida uma descrição (correta ou aproximadamente correta) do mundo em termos de indivíduos nomeados e digitados, e suas relações. Essas pressuposições são críticas para a abordagem da Boa e Velha Robótica.

Antes de irmos adiante com esta descrição, devemos reconsiderar: por que a noção de um sistema de símbolos internos deve estar em contraste com a teoria do *Umwelt*? Talvez trate-se de uma questão absurda, já que as duas teorias parecem ser completamente incomensuráveis. Pode-se no entanto interpretar a situação como se a robótica em IA de fato fosse uma hipótese sobre a

estrutura do *Umwelt* específico humano, que é, de alguma forma e em alguma extensão, simbólica e racional. Mas isto negligencia, primeiro, o fato de que a teoria provê uma epistemologia à parte para o *Umwelt* específico humano no nível da antroposemiose (T. Von Uexküll 1986a, 1986b, 1989) que não pode ser reduzida à hipótese dos símbolos físicos, e segundo, que seu paralelo filosófico é uma versão materialista do funcionalismo em filosofia da mente — a tese de que a mente está para o cérebro como o *software* está para o *hardware*. Evidentemente esta noção é difícil de se compatibilizar com a teoria do *Umwelt*. Ao contrário, os estudos de Jakob Von Uexküll sobre o *Umwelt* específico das espécies de vários animais pode ser visto como antecipando 'estudos ecológicos da percepção' (escola de Gibson) e noções de incorporação e imersividade em PSA (e.g. Hendriks-Jansen 1996) desenvolvidos em oposição à robótica tradicional em IA.

Veículos biomecânicos como sistemas proto-autônomos

As pessoas muitas vezes sonham com a construção de análogos mecânicos precisos de seres vivos, mesmo não sendo considerados muito 'inteligentes'. Considerações históricas podem fornecer exemplos interessantes.[23] Em 1950, W. Grey Walter, o diretor do departamento de fisiologia no Instituto de Neurologia Burdon, em Bristol, publicou o artigo 'Animação da vida'[24] descrevendo duas tartarugas mecânicas, Elmer e Elsie, cada uma delas equipada com dois órgãos sensoriais, e duas células nervosas eletrônicas. Ele as chamou de *Machina Speculatrix*, para ilustrar seu comportamento 'exploratório, especulativo'. Historicamente elas representam exemplos iniciais de 'animats' ou agentes autônomos, construídos a partir de princípios cibernéticos simples.

Cada máquina possui somente duas unidades funcionais, ou sistemas de controle, uma sensível à luz e outra sensível ao toque. Com estes dois órgãos sensoriais (um 'olho', ou fotocélula que pode varrer o ambiente por estímulos de luz, e um simples sensor de toque), dois tubos miniaturas de rádio, dois atuadores ou motores (um para rastejar e um para girar), e uma fonte de energia (baterias), as máquinas podem produzir um comportamento 'natural'. Na ausência de estímulo adequado de luz, Elmer e/ou Elsie exploram o ambiente continuamente (através de fotocélulas ligadas a um mecanismo de giro), com o motor deslocando a máquina para

frente. Os dois movimentos são combinados para dar à máquina um modo de andar cíclico, enquanto a fotocélula 'trava' em cada direção por sua vez. O resultado é que, no escuro, Elmer explora uma área considerável, permanecendo alerta à possibilidade de luz e evitando obstáculos que não pode forçar para fora de sua trajetória. Quando a fotocélula vê uma luz, o sinal resultante é aumentado pelos tubos no amplificador. Se a luz é fraca, somente uma mudança de iluminação é transmitida como sinal efetivo. O efeito é deixar o mecanismo de giro de forma que a máquina se move em direção a fonte de luz — analogamente ao comportamento biológico conhecido como 'tropismo positivo' (por exemplo, uma mariposa voando para uma vela).

Mas Elmer não se desloca para dentro da fonte de luz: quando o brilho excede um certo valor, o sinal se torna forte o suficiente para operar um relê no primeiro tubo, que tem o efeito inverso do segundo. O mecanismo de giro é ligado novamente ao dobro da velocidade de modo que a máquina desvia e procura um clima mais ameno. Ela circula ao redor de uma única fonte de luz, segundo um caminho complexo de avançar e se afastar; com duas fontes de luz ela continuamente passeia entre elas. Quando as baterias estão bem carregadas, ela é atraída para a luz distante, mas no limiar o brilho é forte o suficiente para agir como repelente de forma que a máquina se desvia para continuar a exploração. Quando o nível das baterias enfraquece, a sensibilidade do amplificador é aumentada de forma que a atração pela luz é sentida de mais longe. Mas assim que o nível de sensibilidade cai, a maquina eventualmente está na entrada de sua 'toca' (uma caixa que emite luz com um certo brilho) e ela é atraída direto para casa, pois a luz não parece mais tão ofuscante. Na 'caixa toca', ela faz contato com o carregador e suas baterias são recarregadas.

Walter experimentou variações desta configuração e observou como novos comportamentos complexos podiam emergir em interações de duas máquinas se elas pudessem se avistar (quando uma pequena luz é montada na casca das tartarugas). Ele notou que essas máquinas, apesar de grosseiras, forneciam 'uma estranha impressão de propósito, independência e espontaneidade' (1950: 45).[25] Aparentemente estes dispositivos se comportavam como se tivessem agenciamento autônomo, e podemos até perguntar, como veremos, se eles têm um *Umwelt* primitivo (apesar de Walter, segun-

do meu conhecimento, nunca ter pensado nisso), como J. Von Uexküll enfatizou para carrapatos, besouros e outras criaturas.

Naquele tempo, estas máquinas (incluindo versões posteriores modificadas por Walter) pareciam modelos poderosos de comportamento autônomo.[26] No entanto, durante a década de 1950 e 1960, esforços mais consideráveis foram feitos na tentativa de construir programas inteligentes que podiam simular capacidades cognitivas de alto-nível, e Walter continuou seu trabalho em outras direções.

Em 1984 é publicado o livro de Valentino Braitenberg — *Vehicles: Experiments in Synthetic Psychology*[27] — que se tornou uma importante referência não apenas para psicólogos com foco cibernético, mas para trabalhos em Vida Artificial. Braitenberg, que tinha formação em cibernética e neuroanatomia, descrevia uma variedade de pequenas e simples criaturas, máquinas com sensores e ação motora (principalmente rodas), projetadas com técnicas básicas de engenharia, que podiam criar uma grande diversidade de formas de comportamento. Braitenberg, que também estava interessado em estruturas cerebrais de animais que pareciam ser 'pedaços de maquinaria computacional', considerou os veículos (eles não usou o termo sistemas autônomos) como se fossem animais em um ambiente natural. Então, alguém pode se sentir tentado a usar a linguagem psicológica na descrição deste comportamento, mesmo que saiba que, de acordo com ele, não existe nada neste veículo que os projetistas não tenham colocado ali (algo similar à 'postura intencional' de Dennett[28]). Braitenberg fez algumas observações que são de importância geral para o desenvolvimento de sistemas autônomos, e nós iremos considerar as mais importantes.

O primeiro ponto é sobre o tipo de física onde o veículo 'vive'. Um veículo deve ser capaz de mover-se, mas até planetas se movem, então o que é especial no movimento de um sistema autônomo? Braitenberg descreve a espécie mais simples — Veículo 1 — como equipado somente com um sensor na frente e um motor atrás com uma conexão muito simples entre sensor e motor (Figura 1).

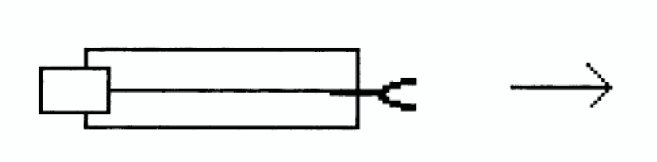


Figura 1. Veículo 1. A caixa a esquerda é o órgão motor; o corpo com a conexão sensório-motor está no meio; e a forquilha em Y é o sensor; a seta indica a direção do movimento.

Quanto maior o nível de uma certa qualidade (por exemplo, temperatura) para a qual o sensor está ajustado, mais rápido o funcionamento do motor. O veículo se move na direção para qual estiver apontando; ele desacelera nas áreas frias e acelera onde está quente. Mas ele vive na terra (ou na água), isto é, em um mundo no qual a lei de Newton de inércia não faz sentido diretamente; é um mundo de fricção, um mundo Aristotélico, neste sentido. A Fricção desacelera o corpo, e se o veículo entra em uma região fria onde a força exercida pelo motor, sendo proporcional à temperatura, se torna menor que a força de fricção, ele para. Braintenberg agora nos pede para imaginar um veículo deste tipo nadando em um lago: 'Ele é incansável, você diria, e não gosta de água quente. Mas é bastante estúpido, uma vez que não é capaz de voltar para um bom lugar frio que ele errou ao tentar alcançar, em sua inquietação. De qualquer forma, você diria, ele está VIVO, uma vez que você nunca viu uma partícula de matéria morta se mover desta maneira' (Braitenberg 1984: 5).

Por meio de incrementação, Braitenberg aumenta a complexidade da série de veículos. O veículo 2 é um tipo de duplicata do primeiro, com dois motores e dois sensores, nos cantos do chassi, e com duas variedades de conexões sensório-motoras, direta e cruzada (ver Figura 2). Se não há cruzamento, o motor no lado do corpo que recebe maior exposição ao sensor, do que aquele que ativa o sensor, tende a se mover mais rápido, de forma que, como resultado, o veículo se vira para longe da fonte (tem 'medo' dela, como diz Braitenberg). No veículo com cruzamento, o movimento resultante vira o veículo para a fonte (indicada pelo ícone do sol) e, eventualmente, colide com ela.

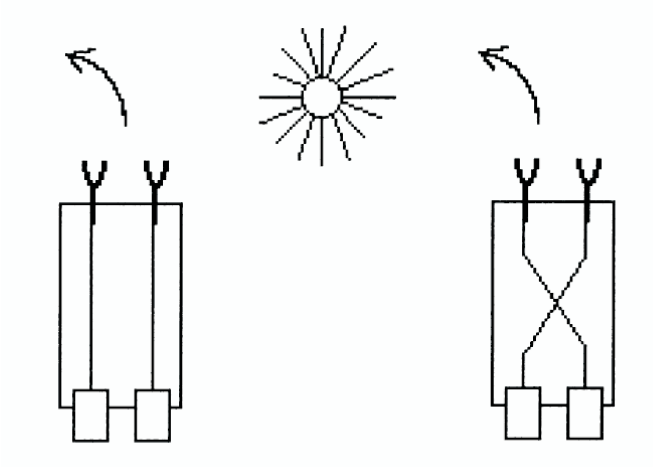


Figura 2: Veículo 2

Isto é somente o começo. Braitenberg desenvolveu uma série de agentes com um amplo espectro de capacidades que ele interpretou como 'medo', 'agressão" etc. Não vou me deter nos detalhes. A importância para PSA é clara. Mesmo que princípios cibernéticos possam servir como base para 'quebrar o problema da cognição' (por exemplo, construir sistemas que poderiam exibir inteligência geral, capacidade de solução de problemas gerais, planejamento, etc.), a construção de veículos biomecânicos revelou que comportamentos de máquinas simples podem, em um ambiente variado, e interagindo com outras máquinas, produzir algo que aparenta ser organismos governados por estruturas de controle quase inteligentes. Como animais vivos, eles parecem consistir em ciclos funcionais simples de processos semióticos de interpretação e ação sígnicas. Por que eles não poderiam ter *Umwelt*s?

As várias tentativas de construir veículos biomecânicos autônomos[29] diferem em uma questão importante dos primeiros autômatos de relojoeiros (por exemplo, o autômato 'desenhista' da família Jaquet-Droz, cf. Chapuis e Droz 1958), como também dos robôs em IA que se seguiram ao período cibernético: eles não dependem de um 'programa' central. Nos robôs e autômatos do século XIX, o programa era responsável pelo comportamento dinâmico do modelo. Fosse um tambor de rotação com lingüetas disparando alavancas em seqüência, um conjunto de câmeras motorizadas, ou outro mecanismo, o movimento do autômato criado era 'ativado' por um maquinário controlador central. Como enfatizado pelo novo movimento de pesquisa em sistemas dinâmicos complexos[30], aí está a

fonte das falhas destes modelos, e a perspectiva limitada de um programa completo para modelar sistemas humanos e animais que se desenvolviam e vieram a incluir boa parte da IA. As abordagens mais promissoras para modelar fenômenos complexos, como vida ou inteligência, foram aquelas que dispensaram um controlador global centralizado e focaram suas atenções em mecanismos de controle distribuído do comportamento, atividade situada, e os tipos de dinâmicas emergentes que se formam a partir de agentes locais em interação. Grey Walter e Valentino Braitenberg fizeram um trabalho pioneiro na modelagem de sistemas autônomos sem pressupor estruturas de controle elaboradas e explicitamente codificadas. De uma perspectiva biossemiótica, é uma ironia histórica que ambos, o paradigma relativamente mecânico da cibernética, e a idéia 'vitalista' de *Umwelt*, e sua defesa da noção de unidade do organismo e seu ambiente sensorial, devem ser vistos como precursores dos conceitos mais recentes de cognição situada e incorporada desenvolvida no contexto de PSA. Vejamos esta pesquisa.

Agentes autônomos

Durante a década de 90, conceitos e métodos foram usados para revitalizar o melhor da antiga abordagem cibernética para execução de projetos de robôs simples (revivido em Braitenberg) em um conjunto que se tornou conhecido como projeto de sistemas, agentes ou animats autônomos. Já em meados da década de 80, Rodney A. Brooks, e seu grupo do Laboratório de Inteligência Artificial do *Massachusetts Institute of Technology*, desenvolveram uma crítica consistente do 'paradigma de Pensamento Deliberativo'[31], e da tese de que tarefas inteligentes podem (e devem ser) implementadas por processos de raciocínio operando em um modelo simbólico interno. Essa crítica, mais um novo conjunto de técnicas de modelagem e princípios de construção, gradualmente ganharam em influência[32] e se tornaram conhecidas como o 'movimento dos sistemas reativos' e 'pesquisa em agentes'. De forma paralela, próxima deste movimento e profundamente inspirada por ele, novas noções de processos cognitivos como sendo executados em sistemas incorporados e situados foram desenvolvidos.[33] De um modo interessante, uma das fontes na qual este novo paradigma de construção de robôs e cognição incorporada foi inspirou-se foi a teoria do *Umwelt* uma vez que ela foi usada para enfatizar, primeiro, a forte conexão dinâmica entre o corpo do animal e seu mundo de experiênci-

as, e segundo, que o mundo percebido pelo animal é diferente do mundo percebido pelo biólogo, indicando a necessidade de maior atenção para o fato de que um robô pode viver em um mundo 'perceptivo' que difere daquele que o construtor do robô pode ver de forma imediata.[34]

Os princípios de projeto de agentes

O grupo de Brooks, e outros pesquisadores, acharam irrealista esperar que tarefas orientadas para ações podessem ser implementadas de forma bem sucedida em uma máquina 'deliberativa', em tempo real. Eles então começaram a desenvolver novas idéias sobre como agentes autônomos deveriam ser organizados, um projeto que, de acordo com Maes (1990), levou a arquiteturas radicalmente diferentes. Estas arquiteturas (e.g. Maes 1990; Brooks 1986, 1991b; Brooks 1992; Meyer e Guillot 1991; Wilson 1991; Brooks & Mães 1994; Clark 1997; Ziemke e Sharkey 1998) são geralmente caracterizadas por:

- funcionalidade emergente
- decomposição baseada em tarefas
- acoplamento mais direto entre percepção e ação
- distribuição e descentralização
- interação dinâmica com o ambiente
- fundamentação física (imersividade e corporificação)
- mecanismos intrínsecos para lidar com limitações de recursos e conhecimento incompleto

A *funcionalidade* de um agente é considerada uma propriedade emergente da interação do sistema com seu ambiente dinâmico.[35] A especificação do comportamento de um agente não explica a funcionalidade que é exibida quando ele está operando. O que parece ser um comportamento complexo não precisa necessariamente estar codificado no agente, podendo ser o resultado de poucas, e simples, regras comportamentais e interação com o ambiente. O ambiente não é tratado preditivamente, mas sua caracterização é explorada para servir à funcionalidade do sistema. Assim, não se pode dizer a estes agentes como atingir um objetivo. É preciso achar o laço de interação envolvendo o sistema e o ambiente (dado que o ambiente possui propriedades esperadas) em direção ao ob-

jetivo desejado (isso pode parecer simples, mas de fato muitas vezes se mostra difícil 'achar' tal laço).

A segunda característica é a *decomposição no nível de tarefas*. Isto não significa o mesmo que decomposição na IA clássica. Um agente é visto como uma *coleção de módulos* cada qual tendo seu próprio domínio específico de interação, ou competência. Os módulos operam quase-autonomamente e são responsáveis por sensoriar, modelar, computar, raciocinar, e controlar o que é necessário para obter uma competência específica. O projeto do agente não se abstém de usar noções representacionais, ou técnicas de raciocínio de IA, mas a moldura conceitual em que estas noções são desenvolvidas mudou, porque não existe módulo central de raciocínio que planeja e governa o comportamento completo, nem qualquer atividade global de planejamento em uma estrutura hierárquica de objetivos. Para evitar duplicações caras e desnecessárias de módulos, eles podem fazer uso de 'sensores virtuais'. A comunicação entre os módulos é reduzida ao mínimo e não opera por meio de linguagens de alto nível, mas em um nível de baixa informação. O comportamento geral do agente não é uma composição linear dos comportamentos de seus módulos, mas emerge através de interações com os comportamentos gerados pelos módulos.

O *acoplamento direto da percepção e ação* é facilitado pelo uso de métodos de raciocínio que operam em representações que estão próximas da informação dos sensores (representações 'análogas'[36]). Se um problema como categorização de objetos pode ser resolvido através de um processo que lida com sensação ou percepção, ao invés de operações simbólicas, então ele é preferido. A percepção pode se tornar menos geral, embora mais realista, e não há necessidade do sistema perceptivo enviar uma descrição do mundo como em IA. A especial 'arquitetura de subsunção'[37] permite, ao projetista, conectar mais fortemente percepção a ação, fazendo comportamentos de robôs emergirem concretamente no mundo, para usar outra frase popular nesta abordagem. Mais uma vez, nós podemos perguntar: Porque não ver isto como uma tentativa de desenvolver uma teoria específica das atividades internas de um *Umwelt*? Logo vamos voltar a esta questão.

A abordagem de agentes, ou 'nova IA', baseia-se na *hipótese da fundamentação física*. Ela afirma que para construir um sistema inteligente é necessário que suas representações estejam fundamentadas no mundo físico.[38] O que exatamente isto significa raramente

é descrito de forma explícita, mas algumas pistas podem ajudar. Um sistema fisicamente fundamentado é um sistema que está conectado ao mundo por meio de sensores e atuadores. Desta forma não é adequado estudar, por exemplo, problemas de percepção por técnicas simples de simulação; entradas e saídas digitadas não são mais de interesse porque não estão fisicamente fundamentadas. Está embutida aqui a idéia de que sistemas devem ser construídos de baixo-para-cima. Abstrações de alto-nível precisam se tornar concretas. O sistema construído deve expressar todos seus objetivos como ação física (em oposição a representações não-dinâmicas armazenadas na memória), e o sistema deve extrair todas suas informações dos sensores físicos, isto é, a 'entrada' não deve ser entregue aos sistemas como informação simbólica, mas como ação física. O projetista de tal sistema é forçado a fazer mais componentes explícitos. Cada 'atalho' tem impacto direto sobre a competência do sistema; não existem ligações frouxas na conexão entrada/saída. A própria noção de representação como algo explícito e estável é criticada.[39] Isto levou pesquisadores ao que é chamado 'visão antirepresentacionalista de cognição', que, no entanto, é uma maneira inadequada de expressar o fato de que a *Velha e Boa IA* tem uma visão simplista de categorias como 'representação' e 'símbolo'. Deve-se reconstruir diversos tipos de representação em vários tipos de sistemas como um contínuo de casos dentro de um modelo semiótico geral de representação, como sugerido por Katz & Queiroz (1999). Outro modo de definir a idéia de fundamentação física é por meio das noções de imersão e corporificação (Brooks 1991a, 1991b; cf. Hendriks-Jansen 1996). *Imersão* indica que os robôs estão situados em um mundo. Eles não tratam de descrições abstratas, mas do aqui-e-agora do ambiente que diretamente influencia seus comportamentos. *Corporificação* indica que o robô tem corpo, experimenta o mundo diretamente e que as ações tem realimentação imediata sobre o equipamento sensório do próprio robô.[40] Robôs simulados em computador podem até estar 'imersos' (situados) em um ambiente virtual, mas eles certamente não estão corporificados.

Vida e inteligência: as perspectivas da pesquisa sobre agentes

Pode-se perguntar, é claro, se estes requisitos são suficientes para assegurar que um sistema construído com tal e tal comportamento é inteligente (no sentido racionalista de Newell de 'inteligên-

cia geral'). Provavelmente não. Mas aqui pode-se notar uma diferença crucial entre a *Velha e Boa Robótica* e a nova abordagem com respeito ao conceito de inteligência. Para uma abordagem clássica, comportamento inteligente pressupõe capacidade de manipulação racional de elaboradas estruturas simbólicas internas — uma 'linguagem do pensamento' de algum tipo — que representam estados de coisas no mundo real. Embora a 'linguagem do pensamento' não precise ser usada para a comunicação lingüística, sabe-se que poucas espécies (se alguma) têm capacidade representacional da mesma ordem de magnitude e complexidade que o *Homo sapiens*. Por outro lado, pesquisadores concordam que muitos animais são 'inteligentes' de alguma forma. A evolução de animais inteligentes é considerado um padrão instrutivo para entender os requisitos do comportamento inteligente. Computacionalmente, as coisas mais difíceis de se obter, por evolução, parecem ser a habilidade de se mover em um ambiente dinâmico, e processar informação sensória de forma adaptativa para garantir sobrevivência e reprodução. A evolução concentrou seu tempo nesta parte da inteligência, em processos fisicamente fundamentados dos sistemas animais.[41] Da perspectiva da teoria do *Umwelt*, nós podemos ver estas partes como proximamente relacionadas com a emergência de *Umwelt*en complexos. Então, a evolução primária do *Umwelt* é computacionalmente 'custosa': ela toma muitos passos do tempo evolucionário. Este também é o caso de formas de vida simples como células eucariotes unicelulares (Protozoa) que não têm um sistema nervoso e um *Umwelt* genuíno, mas têm (de acordo com T. von Uexküll 1986a) um 'ciclo autocinético ou auto-móvel' simples através do qual entram em interações semióticas com seu meio exterior.

A perspectiva evolucionária da 'nova IA' parece promissora. Um número crescente de especialistas em IA reconhece as limitações da abordagem puramente lógica para construção de 'máquinas que pensam', e são atraídos por princípios biologicamente inspirados, que podem formar a base da arquitetura de *hardware* e *software* em computadores do futuro.[42] Técnicas de Vida Artificial servem de inspiração para encontrar 'modos mais naturais' de observar diversos problemas em projetos de robótica. Os organismos não foram abandonados na natureza após terem sido construídos como projetos funcionalmente perfeitos. A evolução operou como um mecânico que conserta uma máquina quebrada usando o que tem em mãos. Nem todo projeto é um bom projeto. Há muitas tentativas

mas poucos tipos básicos sobrevivem. Os construtores de robôs podem aprender algo estudando o jogo evolutivo. Ao invés de construir máquinas caras e complicadas projetadas para um número limitado de tarefas pré-definidas, pode-se construir, seguindo o conselho de Brooks, uma multidão de máquinas baratas, simples, quase imprevisíveis, e permitir que evoluam gradualmente.

Para a ciência cognitiva, a nova PSA pode levar a uma considerável mudança de perspectiva. Talvez não se possa 'quebrar o problema da cognição' ou criar uma teoria científica compreensível até entendermos o que significa dizer que algo está vivo. A vida veio antes da inteligência real; sistemas autônomos e Vida Artificial devem vir antes da IA. O problema com a pesquisa em IA é que pode ser que ela tenha pulado diretamente para o exemplo mais complexo de inteligência, a inteligência humana. É bastante tentadora a suspeita de que fomos trapaceados porquê computadores fazem coisas que achamos difíceis de fazer. PSA, VA e IA estão em um contínuo de projetos que tentam modelar habilidades cognitivas, adaptativas e de aprendizagem em todos os graus de complexidade que conhecemos. A PSA pode ser vista como uma ciência que trata do nível mínimo de pensamento, o limite mais baixo de manipulação sígnica e de computação: quão simples deve ser um sistema físico antes do qual não pode ser chamado de computacional e vivo (Emmeche 1994b)? Ou, dito em termos de *Umwelt*: qual é o sistema mínimo (artificial ou natural) que realiza seu próprio *Umwelt*?

A 'tese de que a IA deve ser SA'[43] (ou: 'inteligência exige autonomia') pode ser formulada assim: 'A coisa inteligente mais tola que você pode fazer é ficar vivo' (Belew 1991). Animais fazem isso, e animais humanos também fazem. As preocupações são as diferentes maneiras de gerenciar os requisitos de auto-manutenção e adaptação. O comportamento coerente dos organismos pode muitas vezes ser explicado por interações bastante simples com um ambiente rico e variado. Muito da complexidade parece estar no meio. Pense em uma formiga.[44] Ela anda pelo solo da floresta, cuidadosamente evitando grandes barreiras, mas precisa fazer pequenos desvios para achar espaço para arrastar para casa uma agulha de pinheiro. A formiga dá uma pausa em seu trabalho e troca informação com uma companheira. Ela normalmente tem uma rota complexa. Mas a formiga, como sistema comportamental, assim como seu *Umwelt*, é bastante simples. A complexidade é em grande parte reflexo do ambiente no qual ela se encontra. A questão

aqui é que se temos visões sobre como construir robôs sociais, prestativos, ou coisas assim, devemos primeiro descobrir os procedimentos mínimos que permitem a um animal lidar com sua vizinha mais próxima. Isto não parece muito, mas é! Uma formiga nunca pode imaginar o que vai encontrar no caminho. Abertura, adaptabilidade e flexibilidade se tornam mais importantes do que uma resposta imediata pra cada situação concebível, não importando se a resposta pode ser codificada como um quadro, um esquema, um *script*, ou uma das outras técnicas de IA de representar conhecimento. Portanto parece que *Umwelt*, autonomia, ação 'inteligente' e conhecimento corporificado estão acoplados. Mas é esta toda a história?

ALGUÉM EM CASA? PODEM *UMWELTS* SEREM ARTIFICIAIS?

Quando perguntado se é possível para animats ou sistemas autônomos artificiais (construídos por humanos), como robôs, terem *Umwelt*, as pessoas parecem ter duas intuições diferentes. Uma delas é assim expressa: 'Sim, por que não? Se um sistema vivo tão simples como um carrapato tem um *Umwelt*, por que não um robô?' Parece uma resposta bastante razoável. Oposta a esta é: 'Não, claro que não! Que tolo! É somente eletrônica. Não importa quão complicado os circuitos de suas redes neurais artificiais sejam (ou que módulos intermediários entre sensores e motores existam). Como se pode pensar que ele pode sentir algo?' Eu, intuitivamente, tenderia para a resposta 'nenhum-*Umwelt*-em-um-robô', apesar de concordar com alguns argumentos para a resposta 'sim-existe'. Mas como as intuições dividem as pessoas, e podem enganá-las, vejamos alguns argumentos.

A resposta 'robôs tem *Umwelt*' pode ser colocada assim. Premissas: 1. Tudo que é necessário para constituir um *Umwelt* no sentido de um mundo experimentado por fenômenos, específico da espécie (ou 'específico do dispositivo'), é uma certa relação circular baseada em processamento de informação entre dispositivos sensores e dispositivos motores, como descrito pela noção de ciclo funcional. 2. Mesmo criaturas artificiais simples (como Elmer de Grey Walter) instanciam tal ciclo, como animais simples o fazem. 3. Conclusão: sistemas autônomos artificiais como robôs possuem um *Umwelt* (a

partir do que segue por definição que, para o robô, existe algo que é sentido como, ou experimentado como).

A resposta nenhum-*Umwelt*-em-um-robô, ao reconhecer que o robô realmente instancia um ciclo funcional no sentido de um laço causal de realimentação, não crê que este ciclo seja um exemplo real de um ciclo funcional no sentido semiótico de que forma, pela ação do signo, a espinha dorsal de um *Umwelt* experienciado. Por que não? Porque segundo esta perspectiva, o que dá ao *Umwelt* sua característica fenomenal não é o aspecto funcional-cibernético dentro do sistema (e na interface sistema-ambiente), mas o fato de que o organismo vivo é, antes de qualquer coisa, constituído como um sujeito ativo com algum agenciamento. Portanto, somente seres vivos genuínos (organismos e animais) podem viver experiencialmente em um *Umwelt*.

O contra-argumento aqui é que 'nenhum-*Umwelt*' pressupõe o que deveria demonstrar, ao colocar o critério de existência de *Umwelt* no agente como um tipo de capacidade escondida (oculta!), e somente encontrado acidentalmente em alguns dispositivos (orgânicos não artificiais), ao invés de permitir um critério comportamental objetivamente acessível para a existência de um *Umwelt* (por exemplo, a existência daquilo que deve ser descrito como processamento de informação dentro de algum tipo de arquitetura funcional). Portanto, a resposta 'nenhum-*Umwelt*' não é mais do que a reafirmação de uma intuição.

De certa forma este contra-argumento é justo, mas a pressuposição de que somente critérios comportamentais objetivamente acessíveis contam como critério para qualquer coisa que podemos identificar e estudar cientificamente é uma pressuposição externalista que não é verdade para o entendimento de uma enorme gama de fenômenos (fenômenos intencionais, *qualia*, consciência, etc). Isso é assim, ao menos do ponto de vista de algumas tradições científica que não são exclusivamente externalistas (por exemplo, semiótica, fenomenologia, hermenêutica). Além disso, a resposta '*Umwelt*-no-robô' pressupõe que descrever informacional e ciberneticamente a dinâmica do dispositivo é trivialmente o mesmo que identificar e explicar a existência do *Umwelt* de um robô como um fenômeno intrínseco. Dificilmente isso é convincente, uma vez que implica que dispositivos cibernéticos ainda mais simples devem ter *Umwelt*.

Antes de resolver esta questão precisamos examinar: (a) o sujeito do *Umwelt* de acordo com o próprio conceito de *Umwelt*; (b) o aspecto semiótico do *Umwelt* e sua dependência de aspectos qualitativos da ação e interpretação do signo; (c) possível 'implementabilidade' da ação do signo em meio não-orgânico; (d) não-acessibilidade epistemológica do *Umwelt* (ao menos os aspectos qualitativos dele) por quaisquer outros além de seu dono. Finalmente, (e) vamos discutir vários tipos de 'imersão' em PSA e a artificialidade da imersão de robôs.

(a) Se o que significa ter *Umwelt* é ser um sujeito ativo com agenciamento, devemos ter em mente que a maneira como um *Umwelt* existe é ontologicamente diferente da maneira como o ambiente físico existe, ou a maneira como um sistema neural, uma rede biofísica complexa dinâmica, como estudada em neurobiologia, existe, ou ainda a maneira como existe o comportamento observável de um animal, como estudado pela etologia. Dizer que é subjetivo significa dizer que existe ao modo de um sujeito que experimenta ativamente, que não é algo que pode ser visto ou descrito de um ponto de vista externo (cf. T. von Uexküll 1982a, Nagel 1986, Searle 1992).

(b) Que este aspecto subjetivo da sensação, percepção e cognição animal seja acessível à descrição semiótica deve-se à sua característica de basear-se em modelos de relações triádicas, que são não somente adaptativos (e, portanto, significativas em termos biológicos de uma perspectiva funcionalista, Darwiniana, de sobrevivência) mas verdadeiramente significativos para o animal em questão. Um signo pode ter todo tipo de relações com outros signos e todo tipo de efeitos no processo de interpretação (neste caso, por um organismo). De acordo com a noção de signo de Peirce, signos mais desenvolvidos, como símbolos (e 'argumentos'), incluem signos mais simples ('degenerados'), em que aspectos de secundidade e primeiridade são proeminentes. Isto é, os signos internos mediando os *Merkwelt* e *Wirkwelt* de um *Umwelt* (mediando os órgãos perceptivo e motores/operadores) têm um aspecto qualitativo para eles, um aspecto que é freqüentemente negligenciado tanto por semioticistas quanto por biólogos. Um signo pode ser um *token* de um *type* geral (por exemplo, um padrão percebido pode ser reconhecido

pelo organismo como sendo de um tipo perigoso, como um predador), mas ele tem sempre um aspecto de *tone*, isto é, é qualitativamente sentido de alguma maneira (por exemplo, como desagradável). *Tone/token/type* é uma tríade genuína, em que a propriedade de primeiridade do *tone* é sempre parcialmente escondida dentro da propriedade 'objetiva', mais externa, do signo pertencente a um *type*. Isto corresponde à primeira tricotomia de signos em Peirce, que define a natureza do signo, em si-mesmo, segundo a qual os legisignos são instanciados em sinsignos particulares, e sinsignos envolvem qualisignos (qualidades que são signos). Deve-se lembrar que os diferentes tipos de signos, nas classificações de Peirce, não constituem entidades isoladas distintas mas têm relações internas específicas, como relações de inclusão das categorias superiores de signos nas inferiores.[45] O que é um qualisigno? Somente fenomenologicamente podemos nos aproximar da idéia de qualisigno. Ele é de uma natureza experiencial e é, como Peirce afirma, 'qualquer qualidade na medida em que é um signo', 'por exemplo, uma sensação de 'vermelho' (CP 2.254).[46] Assim, o *Umwelt*, como fenômeno semiótico, inclui qualisignos com qualidades 'tonais'. (Que a semiose é um fenômeno de terceiridade não significa que a primeiridade qualitativa dos signos está ausente.)

(c) Os qualisignos podem ser realizados por dispositivos projetados por humanos? Não creio que seja o caso, mas isso depende. De um ponto de vista Peirceano, pode ser o caso (ao menos potencialmente), dependendo de (i) as capacidades semióticas dos materiais constituintes realizarem a formação de hábitos e sentimentos (*feelings*) e (ii) a organização do próprio dispositivo. Porque os materiais são importantes? Isto não é carbo-bio-chauvinismo? Já argumentei que nas células biológicas os aspectos sígnicos das ações internas não são independentes do meio, isto é, a estrutura-processo da 'informação' na célula pode somente ser realizada pelos materiais biomoleculares altamente específicos da célula (Emmeche 1992). Se um dispositivo, como um robô, puder ter a flexibilidade orgânica especial de um animal que permita a ele instanciar qualquer coisa como uma 'lei da mente' — a tendência para deixar os signos se influenciarem por auto-organização —, então é difícil ver porque tais dispositivos não poderiam ser capazes de realizar semiose genuína (incluindo

qualisignos), e, portanto, não serem simplesmente interpretados como fazendo isto ou aquilo, por um observador externo. (Este é normalmente o caso dos robôs: eles são facilmente interpretados como sendo agentes intencionais. Mas ao tomar esta 'postura intencional' de seus construtores, nada nos diz sobre a existência eventual de seus próprios 'sentimentos'). Se este sistema artificialmente construído realiza a ação e o sentimento vivo do qualisigno, então ele tem mente no sentido de Peirce. Mas ele teria *Umwelt*? Existe alguém aí experimentando algo? Lembrem do escopo extremamente amplo de semiose e de mente, para Peirce. Se o próprio universo é permeado de signos, segundo o pansemiotismo de Peirce (Merrell 1996), este estado de coisas pode não nos ajudar a decidir se um robô está experimentando algo, se ele tem um *Umwelt*. Ele pode ter, podemos imaginar, se o qualisigno, e todas as formas superiores de semiose, se tornarem organizadas de tal forma a tornar possível a emergência deste tipo de unidade e coerência da experência que caracteriza 'um *Umwelt*-como-nós-conhecemos' (o nosso próprio).

(d) Mas como podemos saber? Um *Umwelt* só pode ser diretamente conhecido de dentro. Como a não acessibilidade epistêmica de qualquer *Umwelt*, por qualquer observador além de seu próprio 'dono', implica que, quando encaramos um robô, estamos na mesma situação que estamos quando encaramos um carrapato, uma cobra, um beija-flor, ou um cachorro? Eles têm *Umwelt*, mas como este *Umwelt* realmente é, como ele é sentido, é impossível dizer.[47] Estas duas situações são paralelas mas não são iguais. No caso de qualquer animal vivo com um sistema nervoso central (incluindo um carrapato), podemos ter bastante certeza que eles têm um *Umwelt* de alguma maneira. A visão de máquina da vida e de outras pessoas foi transcendida. Filosoficamente, a única solução para o 'problema das outras mentes' (como podemos ter certeza que outras pessoas têm uma mente?) é dizer pragmaticamente que isto não é um problema, ou dizer: porque eu sei que tenho e eles são semelhantes, por analogia eles devem ter. Esta é a melhor explicação para seu comportamento que é conectado com (e parcialmente explicado por) suas mentes. A solução por analogia é também o que fazemos (muitas vezes de forma menos convincente) com animais.

Nós sabemos que 'existe alguém ali' no organismo do cachorro, apesar da dificuldade de inferir o conteúdo real de seus estados (mais ainda para a cobra, e assim por diante). Mas esta inferência analógica usual é de fato baseada na biologia. O cérebro do pássaro é não somente análogo ao nosso cérebro, ele é homólogo (tem a mesma origem evolucionária).[48] Todos descendemos de criaturas muito simples que possuiam o mesmo tipo de órgão chamado sistema nervoso (SN), incluindo o cérebro. Então, no caso do animal, o 'problema do *Umwelt* de outros animais' é respondido pela combinação de uma explicação externalista de homologia de SN (onde SN tem o papel de condição necessária para um *Umwelt*), e um conhecimento internalista do próprio *Umwelt*, mais a inferência analógica mencionada (suportada pela teoria geral do *Umwelt*). Mas no caso do robô, o 'problema do *Umwelt*' é diferente. Ele não pode ser colocado como um 'problema do *Umwelt* de outra máquna' porque não somos máquinas (cf. Kampis 1994), isto é, não podemos usar uma inferência por analogia, nem podemos apelar para homologia evolutiva. Então mesmo que um robô se comporte como um sistema completamente autônomo, a inferência de que ele tem um *Umwelt* não é garantida por estes argumentos.

(e) Isto significa que devemos distinguir entre 'imersão' verdadeira para animais e imersão artificial para robôs no contexto de PSA e VA? Aqueles que entusiasticamente traçaram paralelos entre 'cognição situada' em robôs e em humanos (e.g. Hendriks-Jansen 1996, Clark 1997) devem se esquecer de algumas qualidades da cognição biológica. Uma possibilidade interessante pode ser lembrada: somente sistemas (robôs ou organismos) que possuem *Umwelt* podem realizar comportamentos como o que vemos em vertebrados superiores e em humanos.[49] Se o *Umwelt* é um fenômeno emergente de alto nível, emergindo de processos sígnicos corporificados no sistema nervoso de um agente situado, uma condição necessária para o desempenho estável do sistema pode ser um conjunto de restrições do *Umwelt* sobre padrões particulares de movimento, em um tipo de 'causação descendente'.[50] Em redes neurais artificiais, assim como em Autômatos Celulares computacionais, toda a eficácia causal da dinâmica do sistema pode ser localizada nas regras de nível inferior de comportamento de entrada e saída dos componentes individuais. En-

tão o que parece ser emergente para um observador pode, de fato, ser emergente somente aos olhos deste contemplador (Cariani 1992). Em constraste, para a percepção e cognição intrinsecamente emergentes em animais e humanos, pode ser o caso de que 'mente sobre a matéria' não seja somente especulação metafísica mas, quando transformada dentro de uma teoria geral de *Umwelt* de níveis dinâmicos de interação e níveis semióticos de interpretação, se torna um princípio de auto-organização e causação descendente em sistemas complexos (cf. Lemke 2000). Esta é a posição da variante biossemiótica do organicismo qualitativo: o fenômeno emergente experimentado como componente do *Umwelt* do sistema tem papel direto na dinâmica comportamental deste sistema vivo. Eles constituem um sistema complexo de interpretação (cf. sistema de modelagem interna de Sebeok) que continuamente molda os movimentos individuais do animal. Aqui, nós somos lembrados da velha idéia Aristotélica de que a alma do animal e o movimento animado genuíno são o mesmo fenômeno (cf. Sheets-Johnston 1998, 2000). Também está aqui implicada uma noção forte de complexidade qualitativa de certos sistemas. De acordo com esta noção, um sistema é qualitativamente complexo se (i) é auto-organizado, (ii) têm um *Umwelt* com um *qualia*, e uma condição para (i) é ter (ii) — o que significa, que para ter a capacidade de auto-organização, e um comportamento gracioso de alto nível, o sistema tem de ter algum aspecto de experiência qualitativa. Isto é, o *Umwelt* tem, de alguma maneira, poderes causais para organizar (por causação descendente) o 'ser' total do sistema, de fazê-lo coerente, de dar a ele sua forma de movimento. A própria noção de um ser qualitativo pode ser interpretada como uma propriedade emergente das partes do sistema e suas dinâmicas, incluindo as interações organismo-ambiente e a trajetória genealógica do sistema ao longo do tempo. O ser pode ser interpretado como um ser interacional, imerso, histórico e emergente, o próprio aspecto de agenciamento de um sistema vivo. Esta noção de complexidade não pode ser reduzida a nenhum aumento linear de uma medida quantitativa.

No reino dos agentes que se movem em um mundo desafiante e dinâmico, sugerimos uma diferença entre robôs artificialmente

imersos e animais verdadeiramente situados com um *Umwelt*. Trata-se de uma diferença entre, de um lado, um agente ser capaz de ver para onde se desloca, de vigiar, e observar os passos de alguém (tudo o que um sistema autônomo precisa 'aprender' se ele deve agir como um agente), e de outro lado, de ser capaz de contemplar algo do mundo experiencial do outro, ver e senti-lo com os 'olhos internos' do outro. O grande postulado (não provado e talvez, em princípio, inverificável) do organicismo qualitativo é que ter um *Umwelt* é uma pré-condição para realmente ser capaz de ter autonomia, em ampla escala, com a graciosidade de movimento que somente animais atingiram. Se os sistemas artificiais são parcialmente 'situados', já que não experimentam um *Umwelt*, existe de fato alguma esperança (em uma abordagem mais próxima da biologia teórica, semiótica, pesquisa em sistemas autônomos e ciência cognitiva) de entendermos mais profundamente sistemas autônomos verdadeiramente situados como sendo um tipo de agente semiótico complexo auto-organizado com propriedades qualitativas emergentes.

AGRADECIMENTOS

Gostaria de agradecer a Jesper Hoffmeyer, Ricardo Gudwin, Kalevi Kull, Winfried Nöth, Stanley Salthe e Tom Ziemke por comentários e críticas a versões iniciais deste artigo.

NOTAS

[1] Quanto à questão do *Umwelt* nesta distinção, se poderia ir até mais longe, diferenciando-se entre o *Umwelt*, em um sentido restrito, como o entorno significativo de uma espécie e o *Innenwelt* como uma versão atual deste entorno para um organismo individual (cf. Anderson et al. 1984:13). Mas esta distinção não é necessária no presente contexto. O termo *Innenwelt* não aparece em J. von Uexküll (1940), nem mesmo no glossário de T. von Uexküll para este texto.

[2] Needham é suficiente como exemplo aqui: ‘Estamos hoje perfeitamente certos (...) que a organização de sistemas vivos é um problema, não o ponto de partida axiomático, da pesquisa biológica. Relações de organização existem, mas elas não são imunes do alcance e entendimento científicos. Por outro lado, as suas leis não

parecem ser redutíveis a leis que governam o comportamento de moléculas em níveis mais baixos de complexidade." (do ensaio de 1937 "Integrative levels: a revaluation of the idea of progress", em: Needham 1943).

[3] Sobre o emergentismo britânico: Beckermann (et al., eds., 1992).

[4] A filosofia da biologia de Kant, em *Kritik der Urteilskraft*, foi uma fonte significativa para Jakob von Uexküll.

[5] Concebendo uma base matemática para a mecânica, Galileo (1564–1642) em *Saggiatore* (1623) elaborou uma distinção recomendada por Democritus (c460–371 ac) entre qualidades inerentes ou produzidas por corpos inorgânicos (forma, tamanho, localização no espaço e tempo, e movimento) e todas as outras qualidades que estão no observador, e não na Natureza (calor, som, sabor etc). Robert Boyle (1627–1691) mais tarde chamou esta demarcação de qualidades primárias e secundárias, uma distinção que foi posteriormente sistematizada por Locke (1632–1704).

[6] Embora não seja adequado inserir pensadores em categorias que não expressam adequadamente suas visões, apenas para dar uma indicação aproximada de possíveis representantes das duas posições, a *corrente principal do organicismo* é freqüentemente expressa em correntes tão heterogêneas quanto o neodarwinismo 'clássico' (E. Mayr,etc.), o darwinismo 'dinâmico', com um enfoque em sistemas auto-organizados e seleção (S. Kauffman, D. Depew, B. Weber), vida artificial e as abordagem em agentes autônomos (C. Langton, R. Brooks etc), sistemas em desenvolvimento (S. Oyama, P. Griffiths, E. Neumann-Held), campo morfodinâmico (B. Goodwin), e conceitos hierárquicos de evolução (S. Gould, N. Eldredge, S. Salthe etc). O *organicismo qualitativo* é representado pela biosemiótica (J. Hoffmeyer, T. Sebeok, J. e T. von Uexküll, K. Kull etc), pela 'abordagem animada' (M. Sheets-Johnstone), pela noção de uma ciência biológica de qualidades (B. Goodwin), e, até certo ponto, por estudos de comunicação animal do ponto de vista de uma 'mente ecológica' (G. Bateson), e até pela 'teoria da atividade', derivada da escola de história cultural soviética (Luria, Vygotsky etc). Os casos da Cibernética de Segunda Ordem (H. Von Foerster, G. Pask etc) e do 'internalismo' (Van de Vijver et al., ed., 1998) são mais difíceis de classificar, mas seriam também provavelmente relacionados ao organicismo qualitativo.

[7] Freqüentemente se assume que, na medida em que estes aspectos subjetivos da vida animal, por exemplo, a dor, podem ser vistos como subservientes à sobrevivência do organismo, eles possuem uma explicação funcional adaptativa dentro de um quadro neodarwinista da evolução por seleção natural. Isso não está correto. A seleção não pode 'ver' a dor de um animal. O animal poderia ser, do ponto de vista da história seletiva, tal como um zumbi insensível, que preservou a mesma relação de entrada-saída funcional entre a detecção de ações inflingidas ao organismo e seu comportamento adaptável em função das ações que causaram eventuais detrimentos. O esquema de explicação neodarwinista corresponde a uma abordagem completamente externalista e não pode dizer nada sobre o mundo experiencial interno do animal. Não há razão para que dispositivos darwinianos processando informação altamente organizada sejam capazes de sentir qualquer coisa.

[8] A acusação de carbono-chauvinismo não poderia ser dirigida a disciplinas tais como a ecologia ou a etologia; ela só se tornou possível como resultado da 'revolução molecular' na biologia depois da descoberta de Watson/Crick.

[9] Para críticas, ver Pattee (1989); Kampis (1991); Cariani (1992); Emmeche (1994b); Moreno (et al. 1997).

[10] Embora idéias semelhantes possam ter sido introduzidas, por exemplo Rasmussen (1992) que utiliza a noção de 'circuito de significado' de John Wheeler para postular que um organismo artificial deve perceber uma realidade de algum tipo. Sebeok (1991) relaciona Wheeler e também a idéia de J.von Uexküll à 'doutrina dos signos'.

[11] Em filosofia moral, o termo é usado nesse sentido (e.g. Mele 1995), o que é uma fonte de potencial confusão.

[12] Um exemplo de um uso biológico é a designação 'sistema nervoso autônomo', ou seja, o sistema de fibras nervosas motoras (eferentes) que dão suporte aos músculos lisos e cardíacos e as glândulas (constituindo os sistemas nervosos simpático e parassimpático), que não é 'controlado pela vontade' (da pessoa autônoma) mas sim auto-governado.

[13] Aqui, o conceito de organização é considerado como: as relações que definem um sistema como uma unidade e determinam a dinâmica das interações e transformações a que o sistema pode se

submeter; a organização de um sistema vivo é considerada como autopoiética.

[14] Para uma história da cibernética, ver Mayr (1970); sobre o pensamento sistêmico, ver Lilienfeld (1978).

[15] Um exemplo é uma máquina onde podemos distinguir quatro partes: um disco rotativo *W*, um controlador *G*, um dispositivo de suprimento de combustível *F*, e um cilindro *C*; esta é similar ao controlador que James Watt inventou para regular a velocidade de rotação em máquinas a vapor, onde a saída (velocidade de rotação) regulava a entrada (o vapor). A máquina é conectada ao mundo externo por uma entrada de energia e a 'carga' é considerada como uma variável, fornecendo peso sobre *W*. O ponto central é que a máquina é 'circular' no sentido que *W* dirige *G*, que por sua vez altera *F*, que alimenta *C* que, por sua vez, dirige *W*. Como isso funciona? Quanto mais aumenta *F*, mais aumenta a velocidade de *C*. Quanto maior for *C*, maior a velocidade de *W*. E, como a realimentaçao é negativa, se é maior a velocidade de *W*, diminui a alimentação de *F*. (Se fosse positiva a realimentação, i.e., se um aumento da velocidade de *W* causasse maior alimentação de *F*, a máquina tenderia a *disparar (runaway)*, operando exponencialmente mais rápido até que alguma parte se quebraria). O exemplo se deve a G.Bateson, que muito se inspirou em princípios cibernéticos em sua tentativa de desenvolver uma 'ecologia mental'. Ele nota (Bateson 1979 [1980: 117]) que nos anos de 1930 quando começou a estudá-los, muitos sistemas auto-corretivos já eram bem conhecidos, como casos individuais, mas os princípios da cibernética permaneciam desconhecidos. Ele lista, dentre os casos individuais, a transformação de Lamarck (1809), o controlador de Watt (final do século XVIII), a visão de seleção natural de Alfred Russel Wallace (1856), a análise de uma máquina a vapor com um controlador, de Clark Maxwell (1868), o *milieu interne*, de Claude Bernard, e a *Sabedoria do Corpo*, de Walter Cannon (1932). Poderíamos adicionar ainda a palestra de Felix Lincke, de 1879, *O Relê Mecânico*, que foi provavelmente a primeira tentativa de esboçar uma teoria de controle por realimentação aplicável tanto a máquinas quanto a animais (cf. Mayr 1970).

[16] Não é meu propósito aqui fazer uma reinterpretação peirceana do ciclo funcional de J. von Uexküll. Thure von Uexküll avançou de alguma forma nessa direção (von Uexküll 1982b; ver também Hoffmeyer 1996).

[17] A noção de 'amplificador de inteligência' é vaga porque assistentes como réguas de cálculo, calculadoras de bolso ou mesmo lápis e papel podem ser vistos como 'amplificadores de inteligência' emboram não sejam inteligentes (Gregory 1981). A argumentação de roboticistas, de que a inteligência pode ser realizada artificialmente, pode ser reformulada desta maneira: na concepção de que estes dispositivos são realmente autônomos, a inteligência deles é 'intrínseca'. Ela não é derivada da inteligência humana ou meramente atribuída ao sistema. Isto, é claro, dá origem à pergunta sobre o que significa ser 'realmente autônomo': simplesmente a capacidade de funcionar por algum tempo sem intervenção humana? A habilidade de se mover e se orientar em um ambiente? Ser capaz de resolver problemas simples? Ser um sistema autopoiético? Ter a capacidade de seguir e viver uma vida própria, se reproduzir, e contribuir para manutenção de uma população?

[18] Um comentarista de IA e Robótica observou, uma vez, que o principal objetivo deste tipo de pesquisa parece ser sintetizar a 'mãe perdida' como uma serviçal cuidadosa. Isto, de fato, é uma outra definição de um sistema autônomo completo: uma serviçal artificial que mantém você funcionando.

[19] Robótica tradicional de IA é revisada em, por exemplo, Gevarter (1985); veja também Pylyshyn, ed. (1987).

[20] Para Newell (1980: 170), que formulou esta hipótese, 'A condição necessária e suficiente para um sistema físico exibir ação inteligente em geral é que ele seja um sistema de símbolos físicos'. 'Um sistema de símbolos físicos é uma 'máquina universal' que é fisicamente realizável; 'qualquer sistema de símbolos razoável é universal (em relação aos limites físicos)' (p.169). Newell define 'universal' com referência a tese de Church (também chamada de tese de Church-Turing). Ele claramente afirma que os avanços em IA (como raciocínio e solução de problemas) em grande parte supera o que foi alcançado em outras tentativas de construir máquinas inteligentes, como o trabalho de construir robôs controlados diretamente por circuitos (p.171). Brooks (1990), em sua crítica à tese paradigmática para robótica, define de forma mais livre: 'A hipótese do sistema de símbolos afirma que a inteligência opera em um sistema de símbolos'.

[21] No entanto, da perspectiva da nova robótica (Pesquisa em Sistemas Autônomos) pode-se perguntar até que ponto teóricos como

Simon, Newell, Fodor, e Pylyshyn realmente se importavam com questões teóricas e práticas realmente difíceis envolvidas na construção de robôs.

[22] Esta é a única alternativa que Brooks (1990) vê em sua crítica à hipótese do sistema de símbolos. Note que este é o importante 'problema de fundamentação do símbolo' (Harnad 1990), veja também a série de artigos de Stevan Harnad referenciado em Hayes et al. (1992).

[23] Um famoso exemplo europeu de tais sistemas 'autônomos' projetados pelo homem é o pato mecânico de Jacque de Vaucanson de 1735; veja Chapuis e Droz (1958).

[24] Walter (1950), e o artigo seguinte 'Uma máquina que aprende' (1951).

[25] Assim, Grey Walter não somente antecipou a noção de agentes autônomos, mas também observou comportamento coletivo emergente muito antes dos trabalhos sobre comportamento coletivo e inteligência de enxame (e.g. Varela e Bourgine 1992).

[26] Grey Walter chegou a pensar que 'poderia ser tecnicamente possível construir processos de auto-reparação e de reprodução nestas máquinas' (1950: 45). A este respeito ele foi otimista demais e não reconheceu o problema fundamental de 'realizar' auto-reprodução biológica (Kampis 1991). Ainda assim, foi Walter quem foi o primeiro a mostrar que dispositivos simples de controle podem produzir comportamento 'natural' com aprendizagem.

[27] É um ensaio clássico sobre a síntese de comportamento complexo a partir de interações de componentes simples.

[28] Ver Dennett (1987). A postura intencional é a idéia de que nós não devemos pensar em nosso vocabulário mental de 'crença','esperança','medo', etc., como realmente representando fenômenos mentais genuínos, mas na verdade somente uma maneira de pensar. É um vocabulário útil para predizer e referir-se a comportamento, mas ele não deve ser encarado literalmente como se referindo a fenômenos reais, intrínsecos, subjetivos, psicológicos; é uma questão de levar a 'postura intencional' para qualquer tipo de sistema autônomo (no sentido intuitivo de autonomia), seja ele um inseto, um robô, ou um ser humano.

[29] *Quase-autônomo* pode ser um termo melhor aqui. Quando observamos demonstrações reais (gravados em vídeo ou 'ao vivo' em

conferências) de várias espécies de agentes situados, ou *animats*, das primeiras versões de Braitenberg às mais recentes, notamos que suas performances não são impressionantes. Um típico lugar para estar 'situado' é em um chão plano com obstáculos suaves formando pratos perpendiculares ao chão e nenhuma rugosidade. E ainda assim, estes pequenos heróis muitas vezes ficam presos em um canto ou enroscados nos sensores protuberantes de um companheiro e então, como se o céu tivesse mandado, uma mão (in)visível do criador desce e os coloca de volta sobre as rodas. A arte da suave educação de robôs não ganhou o reconhecimento que merece.

[30] Por exemplo, Weisbuch (ed. 1991), também Emmeche (1997), para uma revisão.

[31] Cf. o artigo de Pattie Maes 'Designing Autonomous Agents' (1990). As idéias de Brooks sobre a arquitetura de subsunção foi inicialmente ignorada, ou atacada, por pesquisadores em robótica tradicional. Mas a abordagem foi gradualmente aceita, e em 1991 Brooks recebeu o prêmio 'Computers and Thought'. A abordagem é conhecida como 'projeto de agentes', 'teoria da arquitetura de subsunção', 'agentes situados', 'IA nova', e 'IA baseada em comportamentos'. O termo 'sistemas autônomos', apesar de também designar a abordagem de Brooks, é usualmente utilizada em um sentido mais amplo, incluindo a teoria da arquitetura de subsunção. Sobre o desenvolvimento das idéias de Brooks, veja também Levy (1992).

[32] Como uma indicação deste fato: a revista *Robotics*, em 1988, teve seu nome alterado para *Robotics and Autonomous Systems*. Veja também Levy (1992) e os artigos de Brooks.

[33] Winograd e Flores (1986); Varela et al. (1991); Hendriks-Jansen (1996); Clark (1997).

[34] Compare também com o comentário de Andy Clark: 'A similaridade entre os mundos operacionais de Herbert [um dos robôs de Brooks da década de 80] e o carrapato [como descrito por J. von Uexküll] é impressionante: ambos dependem de pistas simples que são específicas para suas necessidades e ambos lucram, não se incomodando em representar outros tipos de detalhes' (Clark 1997: 25). Ainda precisa ser analisado de forma mais precisa, por historiadores da ciência, em que extensão a teoria do *Umwelt* determi-

nou o desenvolvimento conceitual da PSA. Para um uso mais crítico da noção de *Umwelt*, em estudos de PSA, veja Sharkey & Ziemke (1998) e Ziemke e Sharkey (2001).

[35] Maes (1990). Maes não define a noção de emergência, que parece ser dependente do observador (cf. Cariani 1992, Emmeche 1994a).

[36] Compare com Steels (1990) que distingue entre (a) representações categóricas e (b) representações análogas, onde (a) inclui representações tanto simbólicas como sub-simbólicas (isto é, redes do tipo perceptron com categorias codificadas em termos de padrões de ativação sobre uma coleção de unidades), enquanto (b) inclui vários tipos de mapas (por exemplo, para informações sensoriais, um mapa de freqüência, um mapa sonar, um mapa de 'cheiro', um mapa de cor).

[37] Um programa de subsunção é construído em um substrato computacional que é organizado em uma série de camadas incrementais, cada uma (geralmente) conectando percepção e ação. O substrato é uma rede de máquinas de estados finitos. Isto é melhor entendido em contraste com o paradigma da Boa e Velha Robótica segundo a qual o robô primeiro percebe o ambiente, então começa a raciocinar sobre ele, tenta construir um modelo do mundo e estabelece planos para alcançá-los. Somente ao término deste processo o robô age, traduzindo 'cognição' em 'comportamento'. Brooks propõe um acoplamento direto entre ação e percepção, sem o 'gargalo da cognição' das arquiteturas tradicionais. Isto não evidencia uma negligencia a um comportamento de obediência a regras. Mas o agente deve consistir em uma série de módulos, e o comportamento multi-modular do agente irá emergir de uma série de ações envolvidas. Assim, a arquitetura de subsunção consiste em camadas de módulos de comportamento que disparam outros comportamentos quando necessários. Note a estrutura de baixo-para-cima: os comportamentos de nível básico lidam com objetos do mundo na base do momento-a-momento. Comportamentos de baixo nível são determinados por entradas sensoriais nas 'pernas', por exemplo. O nível seguinte pode ser um comportamento de 'andar'; um ainda mais alto pode 'explorar' (Brooks 1992).

[38] 'Uma vez que este compromisso é assumido, a necessidade de representações simbólicas acaba completamente. A observação chave é que o mundo é seu melhor modelo. Ele está sempre atualizado.

Sempre contém cada detalhe do que é preciso conhecer. O truque é sensoriá-lo apropriadamente, e com suficiente freqüencia' (Brooks 1990). Steven Harnad propôe, para resolver o problema de fundamentação do símbolo, a construção de híbridos de sistemas sensório-motores simbólicos e não-simbólicos, próximo da idéia de Brooks de 'fundamentação física' (ver Harnad 1990).

[39] Ver, por exemplo, Peschl (1994: 423) que afirma que 'representação' pode ser melhor caracterizada como o encontro de uma relação/covariância estável entre [o ambiente] 'e algo na representação/corpo do sistema. Isto pode ser obtido por mudanças adaptacionais/construtivas no substrato neural que leva a uma dinâmica corporificada capaz de geração de comportamento funcionalmente adaptado'.

[40] Uma questão negligenciada em PSA é se um robô realmente experimenta sensações (ou qualquer coisa), ou tem um corpo no real sentido de um organismo. Do ponto de vista biológico, a última afirmação é trivialmente falsa uma vez que o robô e um animal são construídos e mantidos por tipos fundamentalmente diferentes de processos. Brooks afirma que robôs reais (em oposição aos simulados em computador) são corporificados. Como veremos, esta afirmação é crucial já que queremos declarar que eles também podem ter um *Umwelt*. Ver também o trabalho de Ziemke e Sharkey (2001).

[41] Este é o argumento de Brooks (1990). Como a evolução do primeiro organismo simples vivo na Terra levou cerca de um bilhão de anos, isto foi um processo lento. (Evidências recentes questionam esta estimativa e sugerem que a aparição das primeiras formas de vida foi um processo muito mais rápido). Outro bilhão de anos se passou antes da aparição das plantas fotosintéticas, e quase há um bilhão e meio de anos (ca. 550 milhões de anos atrás) os primeiros invertebrados chegaram — criar organismos com sistemas de processamento de informação são problemas bem complicados. Então as coisas começaram a se mover rapidamente. Répteis chegaram cerca de 370 milhões de anos atrás, mamíferos 250 milhões de anos atrás, os primeiros primatas apareceram cerca de 120 milhões de anos atrás, os predecessores dos grandes macacos somente 18 milhões de anos atrás. Criaturas como homens chegaram há 2,5 milhões de anos atrás. Os homens inventaram a agricultura há 19 mil anos atrás, e desenvolveram escrita e 'conhecimento especializado' há menos de 5000 anos atrás. Assim, comportamen-

to de solução de problemas, linguagem, conhecimento especializado e raciocínio parecem ser bem simples, uma vez que a essência de ser e reagir estejam disponíveis!

[42] Veja comentário de Belew (1991), que trabalhou com aprendizado de máquina.

[43] Stanley Salthe prontamente destacou que a relação correta não é IA = SA, mas {SA {IA}}.

[44] De fato, muito da pesquisa em Vida Artificial e 'inteligência de enxame' tem se preocupado em entender a estrutura de padrões de comportamento coletivo em formigas, vespas, e outros insetos sociais. Veja, por exemplo, Deneubourgh (et al. 1992), artigos em Kull (2001) e artigos sobre comportamento coletivo em Morán (et al. 1995).

[45] É o que Liszka (1996: 46) chama de 'regra de inclusão'. Esta regra se aplica a cada uma das três principais divisões (isto é, de acordo com a característica do signo em si (qualisigno-signsigno-legisigno); de acordo com a relação com o objeto (ícone, índice, símbolo); com o poder do signo de determinar o interpretante (rema, dicente, argumento)). A implicação lógica é que para cada possível tipo de signo, entre as dez classes, todos incluirão um qualisigno. Mesmo um argumento, que é um legisigno simbólico, deve incluir um qualisigno, embora seu aspecto qualitativo não seja dominante. Então, minha análise segue Liszka (e Peirce), com ênfase especial ao aspecto fenomenal e qualitativo da semiose.

[46] CP 2.254 em Peirce (1931-58) [p. 115 em Peirce 1955]. Nós raramente experienciamos somente qualisignos em nosso *Umwelt*; eles são pensados como o fundo sensível de nossa percepção. Merrell (1996: 43) descreve a percepção de um desenho de um cubo de Necker como oferecendo um 'exemplo desta visão nua, passiva em contraste com ver como é e o que é (...) ver o desenho como a imediaticidade confere um sentimento ou sensação de nada mais que uma qualidade (Primeiridade): Brancura pontuada com 'negridão' fina intermitente. Em um momento mais tarde ela é vista em termos de alguma entidade existente 'por aí' na 'realidade semiótica', como um conjunto de linhas interconectadas. Mas não é (ainda) ativamente visto como um cubo'. Assim, a emergência da percepção do cubo corresponde ao 'desenvolvimento' de um legisigno, que inclui o sinsigno e o qualisigno; no sentido que o qualisigno permeia nosso *Umwelt*.

[47] Apesar de ser realmente possível reconstruir, externamente, um modelo (em nosso *Umwelt*) do mundo fenomenal da criatura; compare Salthe (2001) e Cariani (1996).

[48] Ver Hoffmeyer (2001), referindo-se a Popper, que faz a mesma afirmação.

[49] Uma consideração similar é brevemente feita em uma nota em um famoso ensaio de Thomas Nagel, onde ele enfatiza que o caráter subjetivo da experiência não é analisável em termos de estados funcionais uma vez que estes podem ser designados para robôs que se comportam como pessoas mesmo que não estejam experimentando nada: 'Talvez não possa realmente haver tais robôs. Talvez qualquer coisa complexa o suficiente para se comportar como uma pessoa possa ter experiências.' (Nagel 1974 [1981 p. 392]).

[50] Causação descendente inter-nível não deve ser vista como uma instância da causação eficiente usual (temporal), mas como uma causa funcional e formal. Ver Emmeche, Køppe e Stjernfelt 2000.

Referências

ANDERSON, M.; J. DEELY; M. KRAMPEN; J.RANSDELL; T. A. SEBEOK & T. VON UEXKÜLL. 1984. A semiotic perspectiveon the sciences: Steps toward a new paradigm. *Semiotica* 52(1/2), 7-47.

BATESON, G. 1979. *Mind and Nature*. A Necessary Unity. Wildwood House.

BECKERMANN, A.; H. FLOHR; J. KIM (eds.) 1992. *Emergence orReduction?* Essays on the Prospects of Nonreductive Physicalism. Walter de Gruyter.

BELEW, R. K. 1991. Artificial Life: a constructive lower bound for Artificial Intelligence. *IEEE Expert* 6 (1): 8-15, 53-59.

BRAITENBERG, V. 1984. *Vehicles:* Experiments in Synthetic Psychology. The MIT Press.

BROOKS, R. A. 1990. Elephants don't play chess. Em: *Designing Autonomous Agents:* Theoryand Practice from Biology to Engineering and Back. P. Maes (ed.). The MIT Press. pp. 3-15.

BROOKS, R. A. 1991. New approaches to robotics. *Science* 253: 1227-1232.

BROOKS, R. A. 1992. Artificial life and real robots. Em: *Toward a Practice of Autonomous Systems*. F. J. Varela and P. Bourgine (eds.). The MIT Press. pp. 3-10.

BROOKS, R. A.; MAES, P. (eds.) 1994. *Artificial Life IV*. The MIT Press.

BURKS, A.W. 1975. Logic, biology and automata - some historical reflections. *Int. J. Man-Machine Studies* 7: 297-312.

CARIANI, P. 1992. Emergence and artificial life. Em: *Artificial Life II* (Santa Fe Institute Studies in the Sciences of Complexity vol. X), C.G.Langton, C. Taylor, J. Doyne Farmer e S. Rasmussen (eds.). Addison-Wesley. pp. 775-797.

CHAPUIS, A. & DROZ, E. 1958. *Automata:* A Historical and Technological Study. B.T. Batsford, Ldt.

CLARK, A. 1997. *Being There. Putting Brain, Body, and the WorldTogether Again*. The MIT Press.

DENEUBOURG, J.L.; THERAULAZ, G. & BECKERS, R. 1992. Swarm-made architectures. Em: *Toward a Practice of Autonomous Systems*. F. J. Varela e P. Bourgine (eds.). The MIT Press. pp.123-132.

DENNETT, D.C. 1987. *The Intentional Stance*. The MIT Press.

EMMECHE, C. 1992. Life as an abstract phenomenon: is Artificial Life possible? Em: *Toward a Practice of Autonomous Systems. Proceedings of of the First European Conference on Artificial Life*. F.J. Varela e P.Bourgine (eds.) The MIT Press. pp. 466-474.

__. 1994 a. *The garden in the machine. The emerging science of artificial life*. Princeton University Press.

__. 1994 b. The computational notion of life. *Theoria -Segunda êpoca* 9(21): 1-30.

EMMECHE, C.; KÌPPE, S. & STJERNFELT, F. 1999 (1997). Levels, emergence and three versions of downward causation. Em: *Downward Causation*. P. B. Andersen, N. O. Finnemann, P., V. Christiansen e C.Emmeche (eds.) Aarhus University Press.

GEVARTER, W.B. 1985. *Intelligent Machines:* An Introductory Perspective of Artificial Intelligence and Robotics. Prentice-Hall.

GREGORY, R.L. 1981 (1988). *Mind in Science. A History of Explanationsin Psychology and Physics*. Penguin Books.

HARNAD, S. 1990. The symbol grounding problem. *Physica D* 42: 335-346.

HARRINGTON, A. 1996. Reenchanted Science. *Holism in German culture from Wilhelm II to Hitler*. Princeton University Press.

HAUGELAND, J. 1985. *Artificial Intelligence:* The Very Idea. The MIT Press.

HAYES, P.; HARNAD, S.; PERLIS, D.; BLOCK, N. 1992. Virtual Symposium onVirtual Mind. *Minds and Machines* 2(3): 217-238.

HENDRIKS-JANSEN, H. (1996). *Catching Ourselves in the Act. Situated Activity, Interactive Emergence, Evolution, and Human Thought*. The MIT Press

HEYLIGHEN, F.; ROSSEEL, E.; DEMEYERE, F. (eds.) 1990. *Self-Steering and Cognition in Complex Systems. Toward a New Cybernetics*. Gordon Breach Science Publishers.

HOFFMEYER, J. 1996. *Signs ofMeaning in the Universe*. Indiana University Press.

HOFFMEYER, J. (2001) Seeing virtuality in nature. *Semiotica* 134(1/4), 381-398.

JANLERT, L.E. 1987. Modelling change - the frame problem. Em: *The Robot's Dilemma*. Z.W. Pylyshyn (ed). Ablex Publ.Co. (pp. 1-40)

KAMPIS, G. 1991. *Self-modifying systems in biology and cognitive science:* a new framework for dynamics, information and complexity. Pergamon Press.

KANT, I. 1790. *Kritik der Urteilskraft* (Critique ofJudgement, translated by J. H. Bernard, New York: Hafner PublishingCompany. 1951).

KULL, K. (ed.) 2001. Special Issue: Jakob von Uexküll: A paradigm for biology and semiotics. *Semiotica* 134: 1/4.

LANGTON, C.G. (ed.) 1989. *Artificial Life*. (Santa FeInstitute Studies in the Sciences of Complexity, Vol. VI) Addison-Wesley.

LANGTON, C.G., C. Taylor, J. DoyneFarmer e S. Rasmussen (eds.) 1992. *Artificial Life* II (Santa Fe Institute Studies in the Sciences of Complexity vol. X). Addison-Wesley.

LEMKE, J. 1999. *Opening up closure:* semiotics across scales. (paper delivered at the conf. 'Closure: Emergent organizations and theirdynamics', org. Gertrudis Van de Vijver and Jerry Chandler,Ghent, Belgium, 3-5 May 1999).

LEVY, S. 1992. *Artificial Life. The Quest for a New Creation*. Pantheon Books.

LILIENFELD, R. 1978. *The Rise of Systems Theory*. John Wiley & Sons.

MAES, P. (ed.) 1990. *Designing Autonomous Agents:* Theory and Practice from Biology to Engineering and Back. MIT Press.

MATURANA, H.R. & VARELA, F.J. 1980. *Autopoiesis and Cognition*. Reidel.

MAYR, O. 1970. *The Origins of Feedback Control*. MIT Press.

MELE, A. R. 1995. *Autonomous Agents:* From Self-Control to Autonomy. Oxford University Press.

MERRELL, F. 1996. *Signs Grow:* Semiosis and Life Processes. University of Toronto Press.

MEYER, J.A.; GUILLOT, A. 1991. Simulation of adaptive behavior in aminats: review and prospects. Em: *From Animals to Animats*. J.-A. Meyer e S. Wilson (eds.). The MIT Press. pp. 2-14

MEYSTEL, A. 1998. Multiresolutional *Umwelt*: Toward a semiotics of neurocontrol. *Semiotica* 120 (2/3): 243-380.

MORÁN, F.; MORENO, A.; MERELO, J.J.; CHACÓN, P. (eds.) 1995. *Advances in Artificial Life*. Springer.

MORENO, A.; IBAÑEZ, J. & UMEREZ, J. 1997. Cognition and Life. The Autonomy of Cognition. *Brain & Cognition* 34 (1): 107-129.

NAGEL, T. 1974. What is it like to bea bat. The Philosophical review, October. [reprintedpp. 391-403. Em: *The Mind's I*. D. R. Hofstadter e D. Dennett (eds.) (1981). Penguin Books.

NAGEL, T. 1986. *The view from Nowhere*. Oxford University Press.

NEEDHAM, J. 1943. *Time:* The Refreshing River. George Allen & Unwin.

NEWELL, A. 1980. Physical symbol systems. *Cognitive Science* 4: 135-183.

PATTEE, H.H. 1989. Simulations,realizations, and theories of life. Em: *Artificial Life*. C. G. Langton (ed.). Addison-Wesley. pp. 63-77.

PEIRCE, C. S. 1931-1958. *Collected Papers of Charles Sanders Peirce*. Charles Hartshorne, Paul Weiss e Arthur Burks (eds.), vol.1-8. Harvard University Press.

PEIRCE, C.S. 1955. *Philosophical Writings of Peirce*. (ed. J. Buchler). Dover Publications.

PESCHL, M.F. 1994. Autonomy vs.environmental dependency in neural knowledge representation. Em: *Artificial Life* IV. R. A. Brooks e P. Maes (eds.). The MIT Press. pp. 417-423.

PYLYSHYN, Z.W., (ed.) 1987. *The Robot's Dilemma. The Frame Problem in Artificial Intelligence*. Ablex Publ.Co.

RASMUSSEN, S. 1992. Aspects of information, life, reality and physics. Em: *Artificial Life* II (Santa Fe Institute Studies in the Sciences of Complexity vol.X), LANGTON, C.G., C. Taylor, J. Doyne Farmer e S. Rasmussen (eds.). Addison-Wesley. pp. 767-773.

RICHARDS, R.J. 1987. *Darwin and the emergence of evolutionary theories of mind and behavior*. The University of Chicago Press.

SALTHE, S. N. 2001. Theoretical biology as an anticipatory text: The relevance of Uexküll to current issues in evolutionary systems. *Semiotica* 134(1/4): 359-380.

SANTAELLA BRAGA, L. 1994. Peirce's broad concept of mind. *European Journal for Semiotic Studies* 6 (3/4): 399-411.

SEARLE, John 1992. *The Rediscovery of the Mind*. The MIT Press.

SEBEOK, T. A. 1991. *A Sign is just a Sign*. Indiana University Press.

SHEETS-JOHNSTON, M. 1998.Consciousness: a natural history. Journal of Consciousness Studies 5 (3): 260-294.

SHEETS-JOHNSTON, M. 1999. *The formalnature of emergent organization*. (paper presented at the conference 'Closure: Emergent organizations and their dynamics', Ghent,Belgium, 3-5 May).

STEELS, L. 1990. Exploiting analogical representations. *Robotics and Autonomous Systems* 6: 71-88.

VAN DE VIJVER, G.; S. N. Salthe; M.Delpos (eds.) 1998. *Evolutionary Systems*. Biological andepistemological perspectives on selection and self-organization. Kluwer.

VARELA, F. J.; P. Bourgine, eds. 1992. Toward a Practice ofAutonomous Systems. Em: *Proceedings of the First European Conference on Artificial Life*. The MIT Press.

VON UEXKÜLL, J. 1940. *Bedeutungslehre*. (Bios 10.Johann Ambrosius Barth, Leipzig), [translated by Thure von Uexküll,1982: The theory of meaning. Semiotica 42(1): 25-82; glossery p. 83-87 by T.v.U.].

VON UEXKÜLL, T. 1982a.Introduction: Meaning and science in Jakob von Uexküll's concept of biology. *Semiotica* 42(1): 1-24.

VON UEXKÜLL, T. 1982b. Semiotics and medicine. *Semiotica* 38 (3/4): 205-215.

VON UEXKÜLL, T. 1986a. From indexto icon, a semiotic attempt at interpreting Piaget's developmentaltheory. Em: *Iconicity. Essays on the Nature of Culture. Festschrift for Thomas A.Sebeok on his 65th birthday.* Paul Bouissac, M. Herzfeld e R. Posner (eds.). pp.119-140. Tübingen: Stauffenburg Verlag.

VON UEXKÜLL, T. 1986b. Medicine and semiotics. *Semiotica* 61(3/4):201-217.

VON UEXKÜLL, T. 1989. Jakob von Uexküll's *Umwelt*-Theory. Em: *The Semiotic Web 1988*. Th.A. Sebeok e J. Umiker-Sebeok (eds.). Mouton deGruyter. pp. 129-158.

VON UEXKÜLL, T., Geigges, W. and Hermann, J.M. 1993. Endosemiosis. *Semiotica* 96 (1/2): 5-51.

ZIEMKE & SHARKEY 2001. A stroll through the worlds of robots and animals: applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica* 134 (1): 701-746.

WALTER, W.G. 1950. An imitation oflife. *Scientific American* 182 (5): 42-45.

WALTER, W.G. 1951. A machine that learns. *Scientific American 185* (2): 60-63.

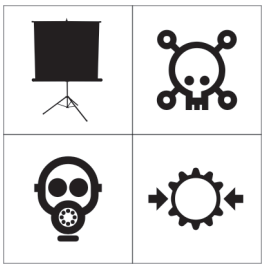
WIENER, N. 1948. *Cybernetics- or control and communication in the animal and the machine*. The MIT Press.

WILSON, S. W. 1991. The animat path to AI. Em: *From Animals to Animats*. J.-A. Meyer & S. Wilson (eds.). The MIT Press. pp. 15-21.

COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA



CAPÍTULO 8

ROBOSEMIÓTICA, COGNIÇÃO ENATIVA E INCORPORADA

Tom Ziemke

INTRODUÇÃO

Muitas pesquisas em ciência cognitiva, particularmente em IA e em vida artificial, têm, desde meados dos anos 80, sido dedicadas ao estudo dos chamados *agentes autônomos*. Tratam-se, especialmente, de sistemas robóticos situados em algum ambiente, com o qual interagem por meio de sensores e efetores. Tais sistemas são freqüentemente auto-organizados, no sentido em que aprendem artificialmente, se desenvolvem e evoluem em interação com seus ambientes, por meio de técnicas de aprendizagem computacional, tais como redes neurais artificiais e algoritmos evolutivos. Devido à motivação e inspiração biológica subjacente a muitas destas pesquisas (Sharkey e Ziemke 1998), agentes autônomos são chamados de 'organismos artificiais', 'animats' (abreviação para 'animais artificiais'), 'vida artificial' ou 'biorobôs'. Esses termos não significam as mesmas coisas. Alguns deles se referem a robôs físicos, enquanto outros a simulações em softwares. Mas todos expressam a visão de

que estes mecanismos são substancialmente diferentes de artefatos convencionais, e que, em alguns níveis, são *como* organismos vivos, no sentido de compartilharem algumas de suas propriedades. Aqui, estes sistemas serão chamados de 'organismos artificiais' ou 'agentes/robôs autônomos'.

Este capítulo está interessado no status semiótico e na relevância dos organismos artificiais. Uma questão que vou abordar é se estes organismos são autônomos, se são capazes de semiose, e até que ponto são capazes. Esta não é uma questão simples, já que semiose é considerada algo que, necessariamente, envolve organismos vivos. Morris (1946), por exemplo, define semiose como 'um processo de signos, que é um processo em que alguma coisa é um signo para algum organismo'. Analogamente, Jakob von Uexkull considerou signos como 'de suma importância em todos os aspectos dos processos vivos' (T. von Uexkull 1992), e fez uma clara distinção entre organismos que, como *sujeitos autônomos*, respondem a signos de acordo com sua própria *energia específic*a, e mecanismos inorgânicos, aos quais falta esta energia e, então, permanecem *heterônomos* (isto será discutido mais tarde).

Mecanismos, é claro, podem estar envolvidos em processos sígnicos, em particular, computadores e softwares. Sebeok afirma (com. pessoal, citada por T.von Uexkull 1982) que 'a característica distintiva de entidades vivas, e de máquinas programadas por humanos, é a semiose'. Aos últimos faltam, entretanto, uma 'semântica primária', isto é, um 'significado intrínsico' (Harnard 1990), ou 'conteúdo para a máquina' (Rylatt et al. 1998). Eles derivam suas semânticas do fato de serem programados, observados e/ou interpretados por humanos. Andersen (et al. 1997) tem argumentado em detalhes que computadores, quando capazes de semiose, caem em algum lugar entre os humanos e os mecanismos convencionais. Mas eles, em última instância, derivam suas 'capacidades' semióticas da interpretação de seus designers e usuários. A principal diferença é que sistemas vivos são autopoiéticos, isto é, se auto mantêm, enquanto as máquinas não são (este ponto será discutido em detalhes). Sua 'tentativa de conclusão' é que

> […] a diferença entre a semiose em humanos e em máquinas pode não residir na natureza particular de qualquer um deles. A diferença pode estar no fato de que a semiose, na máquina, pressupõem a semiose humana, e o surgimento em uma pode ser explicado pelo surgimento na outra (Andersen et al. 1997).

As pesquisas em ciência cognitiva e IA são tradicionalmente dominadas, desde suas origens nos anos 1950s, pela *metáfora computacional* da mente – a visão de que a mente humana funciona como um programa de computador. Isso tem conduzido décadas de pesquisas em IA tradicional à *armadilha internalista* (Sharkey & Jackson 1994), cujo foco em programas computacionais desincorporados, e em representações internas, supõem uma realidade externa pré-dada e refletida (Varela et al. 1991). É esquecida a necessidade de fundamentar e 'ancorar' as representações no mundo que supostamente representam. Para cientistas cognitivos, o uso de agentes situados e incorporados oferece uma alternativa, *bottom-up*, para o estudo do comportamento inteligente em geral, e para a representação interna e uso de signos, em particular.

Organismos artificiais, diferentes de softwares, são equipados com capacidades robóticas sensório-motoras, interagem com seus ambientes, e parecem fazê-lo independentemente da interpretação de usuários externos e de observadores. Mais do que isto, tais sistemas são freqüentemente *auto-organizado*s no sentido de que 'aprendem', se 'desenvolvem' e 'evoluem', em interação com seus ambientes, por meio de técnicas de inteligência computacional e, freqüentemente, por mimese de processos biológicos. Alguns exemplos deste tipo de auto-organização, em robôs adaptativos, serão ilustrados na seção 'Trabalho experimental'. Os processos sígnicos e ciclos funcionais, pelos quais os organismos interagem com seus ambientes, são tipicamente auto-organizados, isto é, são o resultado de adaptação em interação com o ambiente, mais do que o resultado da programação de um designer, e, com freqüência, não são interpretáveis por humanos (Prem 1995). Diferentemente de *softwares*, a gênese desses processos *não* pode ser explicada apenas com referência ao design e à interpretação humanas. Portanto, argumenta-se, agentes autônomos são, pelo menos em teoria, capazes de processamento de uma 'semântica de primeira ordem' (e.g. Harnad 1990, Franklin 1997, Bickhard 1998). Se poderia afirmar que seu interesse semiótico e epistemológico surge porque, diferente de máquinas convencionais, o uso de signos e representações por agentes autônomos é auto-organizado e portanto, como em sistemas vivos, é privado e significativo para os próprios agentes.

Muitos pesquisadores não estabelecem mais uma clara divisão entre animais e robôs autônomos. Prem (1998), por exemplo, refere-

se a ambas as categorias como 'sistemas autônomos incorporados', e não distingue, em sua discussão sobre semiose, entre sistemas vivos e não-vivos. Voltaremos a esta questão na seção ‘RNs recorrentes’, depois de fornecer alguns exemplos de nossos experimentos (seção 2).

TRABALHO EXPERIMENTAL: ROBÔS E REDES NEURAIS RECORRENTES

Diversas pesquisas em *robótica adaptativa* dizem respeito à construção de mecanismos de controle em robôs, ao mapeamento entre sinais sensórios e comandos motores, e ao uso de técnicas adaptativas artificiais evolutivas ou de aprendizagem. Em particular, redes neurais artificiais (RNs) são usadas como 'sistema nervoso artificial', e conectam os receptores do robô aos seus efetores. Os robôs usados neste tipo de pesquisa são, com freqüência, robôs móveis (figura 1, como exemplo). Eles recebem entrada sensória de, por exemplo, sensores de proximidade infravermelhos ou câmeras, e controlam os movimentos de suas rodas através de saídas motoras.

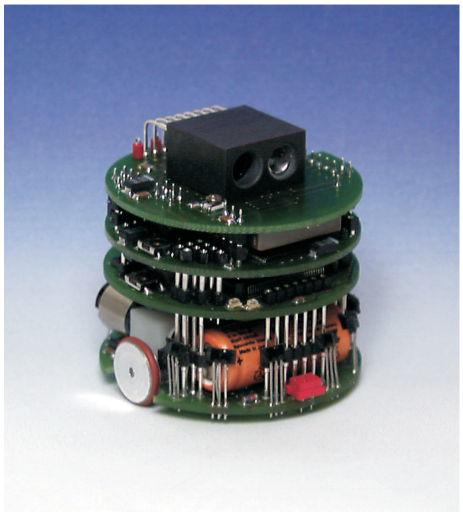


FIGURA 1: o Khepera, um robô miniatura dotado de rodas, freqüentemente usado em pesquisas de robótica adaptativa (manufaturado pela K-Team S.A.; para detalhes, ver: Mondada et al 1993). O modelo mostrado aqui é equipado com 8 sensores infra-vermelhos de curto alcance, ao longo do corpo, e com uma câmera simples, no topo do corpo, de longo alcance.

Um aspecto interessante deste tipo de pesquisa é que ela, em algum nível, oferece um meio para abordar o problema do *símbolo*, ou 'fundamento da representação', em IA (e.g. Searle 1980; Harnad 1990; Sharkey & Jackson 1994; Bickhard 1998; Ziemke 1999). Isso deve-se ao fato de que a rede de controle pode, por meio do corpo do robô (sensores e efetores), interagir com os objetos de seu ambiente, independentemente da interpretação, ou mediação, de um observador. Então, pode-se argumentar que as suas interpretações (signos internos), agora formados a partir da interação física com o mundo que eles 'representam', ou refletem, podem ser considerados fisicamente fundamentados.

Nossa própria pesquisa está interessada no uso de redes neurais recorrentes, que serão discutidas nas próximas seções.

RNs RECORRENTES

O uso de uma rede *feed-forward*, isto é, de uma rede na qual a ativação acontece em apenas uma única direção, de unidades de input para unidades de output, produz um mapeamento input-output que será sempre o mesmo (dado que a rede já aprendeu e não modifica mais o peso de suas conexões). Portanto, o robô controlado será uma 'máquina trivial' (cf. T. von Uexkull 1997), isto é, independentemente da história das entradas (ou do passado), já que toda entrada será mapeada nas mesmas saídas. Em termos semióticos, isto corresponde a uma semiose de informação onde o input corresponde ao signo, o mapeamento input-output corresponde ao interpretante (ou regra causal), e o output ao significado (T. von Uexkull 1997).

Entretanto, se adicionamos feedback interno através de conexões recorrentes, ele se torna uma máquina 'não trivial'. O mapeamento input-output irá variar com o estado interno da rede. A máquina, dependendo de seu passado, pode efetivamente ser uma máquina 'diferente' em cada instante do tempo. Uma analogia, em termos semióticos, poderia ser uma semiose de sintomatização (T. von Uexkull 1997), na qual o interpretante varia e o comportamento input-output do sistema informa a um observador o interpretante corrente. Para o robô, significa que ele não reage meramente a estímulos 'externos', mas 'interpreta' estímulos/signos, na dependência de seu próprio estado interno.

As redes neurais recorrentes têm um importante papel no estudo e na modelagem de representações cognitivas e suas construções. Isto se deve ao fato de que elas explicam: (i) a representação da experiência aprendida (longo prazo), em termos de pesos de conexão, (ii) a representação (curto prazo) do contexto dos agentes controlados, ou o passado imediato na forma de *feedback* interno. Peschl (1997) tem afirmado que RNs, como sistemas nervosos reais, são 'estruturas determinadas' (também, Maturana & Varela 1980). Isto significa que suas reações aos estímulos ambientais dependem do estado corrente do sistema (ou estrutura), não apenas dos inputs. Peschl se referiu a esta propriedade como a '*autonomia* de um sistema representacional'. Ele argumentou que em redes recorrentes o conceito de representação de conhecimento (como 'espelho' da realidade externa) não é aplicável devido ao fato de que não há 'relação representacional estável de referência'. Portanto, o 'objetivo da representação' em tais sistemas não poderia ser a obtenção de um mapa acurado de um ambiente em representações referenciais internas. Ao contrário, sistemas neurais recorrentes deveriam ser vistos como 'dispositivos dinâmicos físicos que incorporam (transformam) o conhecimento para integração sensório-motora [*input-output*], gerando um comportamento adequado, e permitindo a sobrevivência do organismo'. A visão de Peschl de conhecimento, como mecanismos adequados de transformação sensório-motor, torna-se particularmente clara em sua caracterização de conhecimento como 'representação sem representações'.

> As estruturas internas não correspondem às estruturas ambientais; ao invés disso, são responsáveis por gerar funcionalmente um comportamento apropriado que é *desencadeado e modulado* pelo ambiente e *determinado* pela estrutura interna (pesos sinápticos). Isto é o resultado de processos *adaptativos*, filo e ontogenéticos, que alteram a arquitetura ao longo de gerações, e/ou via aprendizagem em um organismo individual, de modo que sua estrutura física incorpora as dinâmicas de manutenção de estados de equilíbrio/ homeostase (Peschl 1997).

Atento às limitações das redes neurais desincorporadas, Peschl sugeriu um conceito de ' 'sistema relativo' de representação como 'determinado não apenas pelo ambiente', mas altamente dependente da 'organização, estrutura, e restrições do sistema de representação bem como dos sistemas sensório-motores que estão incorporados na estrutura de um corpo particular'.

Como já discuti em detalhes (Ziemke 2001), esta visão do papel do conhecimento/representação, em geral, e o uso das redes neurais recorrentes em robôs adaptativos, em particular, é altamente compatível com uma visão enativa de cognição, formulada por Varela (et. al. 1991), e está muito relacionada ao construtivismo radical que Glasersfeld (1995: 51) sumariza assim:

> (i) o conhecimento não é passivamente recebido, ou recebido através de sensores, ou por meio de comunicação;
>
> (ii) o conhecimento é ativamente construído pelo sujeito conhecedor;
>
> (iii) a função da cognição é adaptativa, em um sentido biológico;
>
> (iv) a cognição serve à organização do mundo experiencial do sujeito, não à descoberta de uma realidade ontológica objetiva real.

EXEMPLOS DE TRABALHO EXPERIMENTAL

Nesta seção, irei descrever brevemente dois experimentos de nosso laboratório. O primeiro exemplo é ilustrado na figura 2. Nestes experimentos (Thieme 2002; Thieme e Ziemke 2002; Ziemke e Thieme 2002), robôs (ver figura 1), controlados por diferentes tipos de redes (recorrentes e não recorrentes), são treinados, usando algoritmos evolutivos, para navegar em labirintos de diferente complexidade. No exemplo ilustrado na figura 2, o robô sempre começa na parte inferior esquerda do ambiente e tem de alcançar a área alvo (indicada por um círculo branco). A(s) direção(ões) em que deve(m) manobrar, nas junções (T), são indicadas pelas duas luzes pelas quais ele passa no primeiro corredor em seu caminho para a primeira junção. Neste caso particular, o lado na qual a primeira luz aparece não indica a direção correta na primeira junção, uma vez que o significado da segunda luz é invertido, i.e. ela aparece no lado em direção a qual o robô não deve virar na segunda junção.

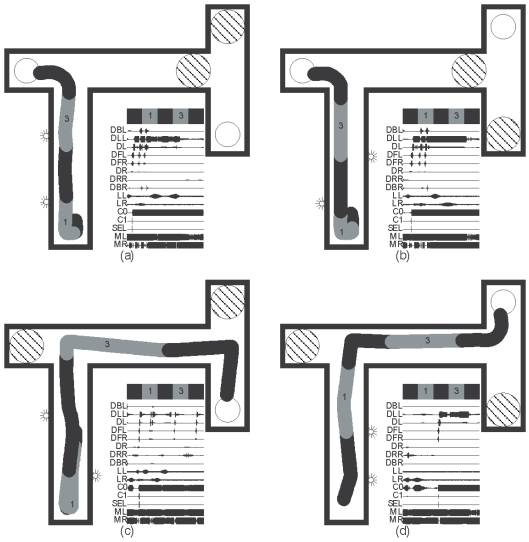


FIGURA 2: Um robô, e suas trajetórias, em 4 exemplos de tarefas de labirinto T. Além disso, as ativações de sensores (D), motores (M) e unidades internas na rede de controle são ilustradas ao longo do tempo. (Adaptado de Thieme 2002; Ziemke & Thieme 2002)

Obviamente, não é fornecido ao robô qualquer informação (acima), e ele tem de achar, por si mesmo, no processo de treinamento evolutivo, um modo de *criar sentido* em sua interação sensório-motora com o ambiente. Em outras palavras, ele deve atribuir significado a diferentes estímulos no contexto. Tendo pleno acesso a todas as coisas que acontecem na rede de controle e no ambiente, podemos analisar em detalhes exatamente como o robô aprende a fazer isto (ver Thieme 2002; Ziemke e Thieme 2002).

O segundo exemplo é ilustrado na Figura 3. Estes experimentos (Buason 2002) investigam a co-evolução competitiva de presas e predadores (de um tipo que está ilustrado na figura 1). A tarefa do predador é caçar a presa, enquanto a tarefa desta é evitar ser caçada. Em uma série de experimentos de crescente complexidade, não apenas os controladores neurais, mas também diferentes aspectos da morfologia dos robôs evoluíram, tais como direção, escopo e ângulo da câmera. Certas coerções são impostas aos robôs, por exemplo: mais poder de visão deve ser 'pago' com redução de velocidade de deslocamento.

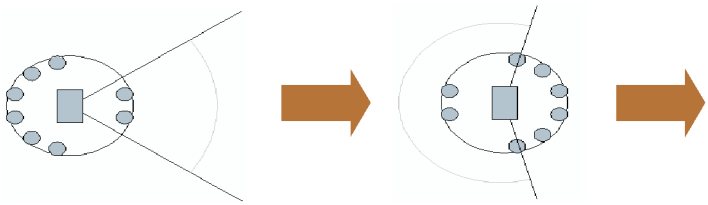


FIGURA 3. Robô-predador (esquerda) e robô-presa (direita), seus campos visuais (indicados por linhas pretas e curvas cinzas) e as direções nas quais se movem. Os pequenos círculos cinzas indicam as posições de sensores infravermelhos de pequeno alcance, seis de um lado e dois do outro.

No exemplo ilustrado na figura 3, o predador desenvolveu um campo visual relativamente estreito (indicado pela luz curva cinza), tendo, conseqüentemente, aumentada sua velocidade. A presa, por sua vez, desenvolveu um campo visual relativamente amplo. Notem que ambos se desenvolveram para que suas câmeras examinem o oponente. Assim, a presa observa a retaguarda, na direção onde há menos sensores infravermelhos, enquanto se move 'para frente'. Além disso, ambos tem suas câmeras apontando na direção onde há menor alcance dos sensores infra-vermelhos, isto é, o predador se desloca de 'costas' e examina a presa na mesma direção. Como no primeiro exemplo, o acesso ao mecanismo neural interno, e ao comportamento dos robôs no ambiente, nos permite analisar em detalhes os processos sígnicos em andamento e sua co-evolução por diversas gerações (ver Buason 2002).

Discussão

Se robôs são autônomos, e se são capazes de semiose, são questões que tem recentemente ocupado um grande número de pesquisadores, em ciência cognitiva e semiótica (e.g. Emmeche 2001; Nöth 2001; Sharkey & Ziemke 1998; Ziemke & Sharkey 2001; Ziemke 2001, 2002; Zlatev 2001). A questão sobre qual é o status semiótico de robôs, não tem, aparentemente, uma resposta simples. A razão disso é que a distinção entre organismos e mecanismos tornou-se difusa, como está indicado em conceitos como 'organismos artificiais' ou 'vida artificial'. Muitos argumentariam que simplesmente não é o caso da gênese da semiose em robôs poder ser (totalmente) explicada com referência a semiose humana, mesmo em se tratando de robôs que se auto-organizam, que se desenvolvem em

interações a longo prazo com seus ambientes, e que são relativamente independentes de seus programadores.

O 'problema' que torna difícil fazer uma clara distinção entre organismos vivos, robôs adaptativos atuais e organismos artificiais, é que os últimos têm, hoje em dia, diversas propriedades do primeiro. Ziemke & Sharkey (2001), por exemplo, discutiram em detalhes as três propriedades que Jacob von Uexküll (1928, 1982) considerou única para organismos (adaptação/crescimento, uso de signos, construção centrífuga), e que podem, em algum nível, ser encontradas nos robôs atuais. Analogamente, Nöth (2001: 695-696) identificou 'quatro razões pelas quais robôs interagem da mesma maneira com seu ambiente como organismos o fazem' e que 'apoiam o argumento de que, não somente organismos, mas também robôs, possuem Umwelt , no sentido de [von] Uexküll: (a) ambos, robôs e organismos, têm Umwelt (ou, na verdade, *Merkwelt*) no sentido que, limitado por sensores disponíveis, eles podem perceber apenas parte de seu ambiente físico; (b) ambos processam estímulos ambientais seletivamente; (c) ambos podem ter 'representações internas de seus Umwelt s'; (d) ambos são equipados com órgãos/módulos perceptuais e orgãos/módulos efetores.

Há, entretanto, um número considerável de diferenças. Corpos de robôs são, por exemplo, de diversas maneiras, extremamente diferentes de corpos vivos, e assim são candidatos improváveis para ter o mesmo tipo de mente fenomenal, ou Umwelt. Em particular, corpos de robôs (*hardware*) e sistemas de controle (*software*) não estão integrados da mesma maneira como estão em corpos vivos. Corpos de robôs, por exemplo, não crescem. Além disso, Ziemke & Sharkey (2001) discutiram em detalhes a falta de *endosemiosis* em robôs e, portanto, a falta daquilo que T. von Uexküll (et al. 1993) refere-se como *neural counterbody*, formado e atualizado em cérebros humanos como resultado de fluxo contínuo de informação de signos proprioceptivos dos músculos, juntas e outras partes de nossos membros, e assim dando origem à experiência do corpo vivo como centro da realidade subjetiva.

Além disso, temos discutido em detalhes (Sharkey & Ziemke 1998, 2001; Ziemke & Sharkey 2001; Ziemke 1999, 2001), que a chave para entender a mente e a semiose pode estar na compreensão de *autonomia e autopoiesis*, isto é, em processos de auto-criação e manutenção, naturais em sistemas vivos (Maturana & Varela 1980). Sistemas autopoiéticos são muito mais integrados do que robôs, no

sentido de que suas ontogêneses começam com uma simples célula a partir da qual eles crescem de uma maneira centrífuga (von Uexküll 1982; cf. Ziemke & Sharkey 2001; Ziemke 2001). Além do mais, eles têm uma intencionalidade natural no sentido de que são unidades autônomas preocupadas com assimilação/ desassimilação do material de/em seu ambiente para o propósito de auto-manutenção e sobrevivência, ao passo que nossos robôs (e.g. presa-predador) só batalham por sobrevivência na visão do observador. Nöth (2001: 696-697) concluiu sua discussão sobre se robôs têm, ou não têm, *Umwelt* como segue:

> É desnecessário dizer que uma máquina, apesar de possuir certa autonomia em seu agenciamento, nunca pode ser dito possuir seu último objetivo em sí. Os objetivos de uma máquina são, sempre, estabelecidos de fora, pelo engenheiro que a projetou e pelo usuário, que a liga e desliga. Assim, o último quadro de referência dos robôs, sua causalidade final, está em outro lugar, e o processo semiótico resultante é aloreferencial.

AGRADECIMENTOS

Gostaria de agradecer a Mikael Thieme e Gunnar Buason pelos experimentos mencionados na Seção 2. Este trabalho foi apoiado pela bolsa #1507/97, da *Knowledge Foundation*, Stockholm.

REFERÊNCIAS

ANDERSEN, Peter B.; HASLE, Per; BRANDT, Per A. 1997. Machine semiosis. Em: Posner, R., Robering, K., and Sebeok, T. A., (eds.), *Semiotik / Semiotics - Ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur / A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, pages 548-571. Walter de Gruyter.

BICKHARD, Mark 1998. Robots and representations. Em: *From animals to animats* 5 - Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior, MIT Press. pp. 58-63.

BUASON, Gunnar 2002. *Competitive Co-Evolution of Sensory-Motor Systems*. Masters dissertation HS-IDA-MD-02-004, Department of Computer Science, University of Skövde, Sweden.

FRANKLIN, Stan 1997. Autonomous agents as embodied AI. *Cybernetics and Systems* 28(6): 499-520.

HARNAD, Stevan 1990. The symbol grounding problem. *Physica D* 42:335-346.

MATURANA, Humberto R. & VARELA, Francisco J. 1980. *Autopoiesis and Cognition - The Realization of the Living*. D. Reidel Publishing.

MONDADA, Francesco; FRANZI, E.; IENNE, P. 1993. Mobile robot miniaturisation: A tool for investigating in control algorithms. Em: *Third International Symposium on Experimental Robotics*, Kyoto, Japan.

MORRIS, Charles W. 1946. *Signs, Language, and Behavior*. Prentice Hall, Englewood Cliffs.

NÖTH, Winfried 2001. Semiosis and the Umwelt of a robot. *Semiotica* 134 (1-4): 695-699.

PESCHL, Markus 1997. *The representational relation between environmental structures and neural systems:* Autonomy and environmental dependency in neural knowledge representation. Nonlinear Dynamics, Psychology and Life Sciences 1(3).

PREM, Erich 1995. Understanding complex systems: What can the speaking lion tell us? Em: Steels, L., editor, *The Biology and Technology of Autonomous Agents*. Springer, Berlin, Germany. NATO ASI Series F, Vol. 144.

PREM, Erich 1998. Semiosis in embodied autonomous systems. Em: *Proceedings of the IEEE International Symposium on Intelligent Control*, Piscataway, NJ. IEEE. pp. 724-729.

RYLATT, Mark; CZARNECKI, Chris A.; ROUTEN, Tom W. 1998. Beyond physical grounding and naive time: Investigations into short-term memory. Em: *From animals to animats* 5 - Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior. The MIT Press. pages 22-31.

SEARLE, John (1980). Minds, brains and programs. *Behavioral and Brain Sciences* 3: 417-457.

SHARKEY, Noel E.; JACKSON, Stuart A. 1994. Three horns of the representational trilemma. Em: Honavar, Vasant e Uhr, L. (eds.), *Symbol Processing and Connectionist Models for Artificial Intelligence and Cognitive Modeling:* Steps towards Integration. NY Academic Press.

SHARKEY, Noel E.; ZIEMKE, Tom 1998. A consideration of the biological and psychological foundations of autonomous robotics. *Connection Science* 10(3- 4): 361-391.

SHARKEY, Noel E.; ZIEMKE, Tom 2001. Mechanistic vs. Phenomenal Embodiment: Can Robot Embodiment Lead to Strong AI?. *Cognitive Systems Research* 2(4): 251-262

THIEME, Mikael 2002. *Intelligence without hesitation*. Masters dissertation HS-IDA-MD-02-001, Department of Computer Science, University of Skövde, Sweden.

VARELA, Francisco J.; THOMPSON, Evan; ROSCH, Eleanor 1991. *The Embodied Mind Cognitive Science and Human Experience*. The MIT Press.

VON GLASERSFELD, Ernst 1995. *Radical Constructivism - A Way of Knowing and Learning*. Falmer Press.

VON UEXKÜLL, Jakob 1928. *Theoretische Biologie*. Springer Verlag.

__. 1982. The Theory of Meaning. *Semiotica* 42 (1): 25-82.

VON UEXKÜLL, Thure 1982. Introduction: Meaning and science in Jakob von Uexküll's concept of biology. *Semiotica* 42 (1): 1-24.

___. 1992. Introduction: The sign theory of Jakob von Uexküll. *Semiotica* 89 (4): 279-315.

__. 1997. Biosemiose. Em: Posner, R., Robering, K., and Sebeok, T. A., editors, *Semiotik / Semiotics - Ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur / A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, Walter de Gruyter. pp. 447-457.

VON UEXKÜLL, Thure; GEIGGES, Werner, & HERRMANN, Jörg M. 1993. Endosemiosis. *Semiotica* 96 (1/2): 5-51.

ZIEMKE, Tom 1999. Rethinking Grounding. Em: Riegler, A.; Peschl, M. and von Stein, A., editors, *Understanding Representation in the Cognitive Sciences*. Plenum Press.

__. 2001. The Construction of 'Reality' in the Robot. *Foundations of Science*, 6(1): 163-233.

__. 2002. On the Epigenesis of Meaning in Robots and Organisms. *Sign Systems Studies* 30(1): 101-111.

ZIEMKE, Tom & SHARKEY, Noel E., (eds) 1998. *Biorobotics*. Special issue of Connection Science 10 (3-4).

__. 2001. A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica*, 134 (1-4): 653-694

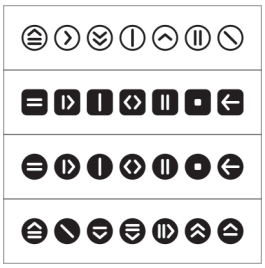
ZIEMKE, Tom & THIEME, Mikael 2002. Neuromodulation of reactive sensorimotor mappings as a short-term memory mechanism in delayed response tasks. *Adaptive Behavior* 10 (3-4): 185-199.

ZLATEV, Jordan 2001. The epigenesis of meaning in human beings, and possibly in robots. *Minds and Machines* 11: 155-195.

COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA



CAPÍTULO 9

FORMA, FUNÇÃO E A MATÉRIA DA EXPERIÊNCIA

Pim Haselager

INTRODUÇÃO

Ao longo da história da Inteligência Artificial (IA)[1], os programadores que criaram softwares capazes de realizar tarefas cognitivamente interessantes, se questionaram: os softwares sabem o que estão fazendo? Eles têm ciência de que estão fazendo alguma coisa? Turing (1950) discutiu este problema sob o título 'o argumento da consciência', e atribuiu sua formulação ao Prof. Jefferson, que abordou a importância de *fazer* como resultado de pensamentos e emoções. Recentemente, parece que este 'problema perene' tem contaminado a *nouvelle* IA (robótica situada e agentes autônomos) através da noção de *Umwelt* (von Uexküll 1936/2001; 1937/2001). Os robôs concebidos são sujeitos que verdadeiramente experimentam suas percepções e ações no mundo? Ou estão mera e artificialmente sinalizando como se estivessem experimentando?

Ao menos algumas vezes os robôs parecem se mover com um propósito, são capazes de evitar certas dificuldades e até parecem

possuir auto-suficiência, evitando danos e redução de energia. Eles parecem saber sobre seus ambientes e são capazes de aprender. Algumas de suas reações a certos eventos parecem baseadas na história de suas interações com o ambiente. A questão é: tudo isto é mera aparência?

É óbvio que diferentes pontos de vista residem nos olhos do observador. Nós temos uma forte tendência para atribuir propósito ou vontade, pensamentos, crenças e/ou desejos, e até mesmo sentimentos a objetos, tais como carros e geladeiras. Coisas que, sob cuidadosa consideração, não seriam qualificadas como possuidoras destas capacidades. No caso de robôs, o perigo de uma interpretação deste tipo é ainda maior. Esta predisposição humana pode alimentar tendências nos cientistas que trabalham com a *nouvelle* IA, como alimentou com os que trabalhavam com GOFAI (Eliza ou MYCIN). O potencial para exploração comercial desta tendência é investigado por companhias que constroem robôs de estimação (e.g. Sony's Aibo and SDR-4X).

Entretanto, o risco de uma sobre-interpretação deste tipo parece não ser maior do que o de uma sub-interpretação. Por exemplo, não serviria a qualquer propósito útil excluir a capacidade que robôs poderiam ter para possuir experiências. Encontramos, nesta direção, argumentos como 'apenas criaturas vivas têm sentimento, propósitos e crenças; robôs não são vivos, portanto, não possuem estas propriedades'. Estes argumentos estão longe de parecer convincentes, não só porque a primeira premissa não está bem estabelecida, como também porque isto é matéria de investigação empírica. É claro que pode ser verdade que robôs não têm sentimentos, propósitos e crenças precisamente porque não são organismos vivos. Só não parece válido partir de premissas desta natureza.

A questão central, à qual vamos dedicar atenção, é se robôs têm, ou podem ter, *Umwelt*. A noção de *Umwelt* foi introduzida por Uexküll (1936/2001; 1937/2001) e indica a experiência subjetiva que um organismo tem de seu mundo perceptual e atuador. A questão está nos aspectos fenomenais de partes específicas do ambiente, motores e perceptuais (Emmeche 2001: 3). Tomamos a noção de *Umwelt*, particularmente relevante para a *nouvelle* IA, porque ela enfatiza a interação de que decorre a experiência: 'estou interagindo com o mundo'. Isto é, a noção destaca mais do que apenas um 'eu' e permite uma abordagem da experiência que não restringe seu foco exclusivamente a aspectos internos.

Nosso argumento é que a noção de vida, como condição necessária para a existência de um '*Umwelt* experimentado', não colabora para que acessemos as capacidades dos robôs. Sugerimos que um olhar atento ao modo como forma, função e matéria interagem pode ser mais produtivo para discutir *Umwelt* em robôs.

O ETERNO PROBLEMA DA IA

Os seres humanos têm sido comparados, ao longo da história, com uma grande variedade de máquinas. No entanto, o mérito de tais comparações tem sido colocado em dúvida desde seus primórdios. Hipócrates (400 a.C.), por exemplo, afirmou: 'Comparar os humanos com seus produtos é expressão de uma visão extraordinariamente empobrecida da humanidade' (*apud* Simmen 1968: 7-8). Mais recentemente, homens foram comparados a relógios (cf. Draaisma 1986). Hobbes (1588-1679) levantou a questão sobre, exatamente, que propriedades deveriam ser atribuídas a relógios.

> A vida não é nada além de movimento de membros. [...] Por que não devemos dizer, que todos os autômatos (máquinas que se movem por molas e rodas, como faz um relógio) têm uma vida artificial? (*apud* Flew 1964: 115)

Descartes (1596-1650) relacionou esta questão com o debate sobre animais:

> [...] não posso compartilhar a opinião de Montaigne, e de outros, que atribuem entendimento, ou pensamento, aos animais. [...] Sei que animais fazem muitas coisas melhor do que fazemos, mas isto não me surpreende. Pode-se usar isto para provar que eles agem natural e mecanicamente, como um relógio, que indica a hora melhor do que faz nosso julgamento. Quando a andorinha vem na primavera, sem dúvida, funciona como um relógio (Descartes, 23 novembro, 1646, carta para o Marquês de Newcastle; Kenny 1970: 206-207).

Baseado na mesma comparação entre relógios e organismos, Descartes se opôs à sugestão de Hobbes. Para Hobbes, a qualidade de auto-movimentação dos relógios levou à questão sobre se poderíamos atribuir propriedades de vida a eles, embora, para Descartes, a similaridade, em certos aspectos (especialmente regularidade), do comportamento dos relógios e dos animais, fornecia motivo suficiente para negar qualquer forma de compreensão aos animais.

De Malebranche (1638-1715) recusou a idéia de que animais pudessem experimentar qualquer coisa:

> Animais não possuem razão ou consciência. Eles comem sem apetite, gritam sem dor, crescem sem compreender; eles não desejam coisa alguma, não temem qualquer coisa, não têm consciência de qualquer coisa (*apud* Wit 1982: 389).

O modo como ele prossegue é especialmente interessante neste contexto:

> Se, algumas vezes, eles se comportam de um modo que parece razoável, então isto é conseqüência de um plano material que Deus ordenou em nome da auto-preservação; eles, sem razão e mecanicamente, escapam de qualquer coisa que ameace destruí-los (*apud* Wit 1982: 389).

Se substituirmos a palavra 'Deus' por 'ser humano' e 'animal' por 'computador', ou 'robô', o resultado é uma sentença que pode ser encontrada nos modelos da IA. Turing, como é bem conhecido, discutiu o 'perene' problema da IA sobre o título 'o argumento da consciência':

> Nenhum mecanismo poderia sentir prazer em seus sucessos (não mera e artificialmente sinalizar, uma fácil maquinação), dor quando a válvula funde, ficar caloroso com bajulações, arrasado com seus erros, encantado com sexo, ficar nervoso ou depressivo quando não consegue algo que quer (Turing 1950: 42).

Basicamente, temos aqui a posição de De Malebranche, aplicado aos computadores, ao invés de aos animais.

O 'eterno' problema tem surgido em diferentes formas. 'Eles' (relógios, computadores, robôs ou animais) não são autônomos, não sabem do que falam suas representações, não são sistemas intencionais, não são capazes de semiose, não possuem originalidade, não são criativos, não tem emoções nem sentimentos, não são conscientes, não estão vivos. Algo que confunde o debate consideravelmente é que qualquer que seja a questão com a qual se comece, rapidamente ela se 'emaranha' em novas questões. Porém, em todas elas está presente a mesma dúvida: há alguém aí?[2]

Nouvelle IA

Os robôs são interessantes candidatos à algumas destas questões, e muitas de suas propriedades parecem dificultar uma resposta completamente negativa. Em primeiro lugar, robôs são criaturas

incorporadas e imersas. Isto é, eles possuem um corpo, diferentemente dos modelos computacionais da IA tradicional, por meio do qual interagem com o ambiente, constituído por objetos e outras criaturas (artificiais e/ou vivas), e sua imersividade no mundo decorre de seu comportamento e de seus processos cognitivos.

Além disso, muitos dos comportamentos dos robôs não parecem pré-determinados, sendo, ao contrário, emergentes. Emergência é um conceito complicado.[3] Vamos enumerar alguns de seus aspectos mais relevantes, neste contexto. Em primeiro lugar emergência pode ser entendida como funcionalidade não programada (Clark 2001: 114). O comportamento de um robô não é diretamente controlado ou programado de um modo simples e direto, mas surge das interações entre um número limitado de componentes que podem ser substancialmente diferentes em suas propriedades e possibilidades de ação. Clark dá o exemplo de simples disposições comportamentais (e.g. tende para a direita, salta de volta quando toca algo) em um robô que, sob certas circunstâncias, poderia prover um comportamento emergente, tal como seguir muros, em protocolos de navegação espacial. Em segundo lugar, um importante aspecto da emergência é que níveis superiores (globais) influenciam, restritivamente, o comportamento e as interações dos componentes em níveis inferiores. Isto, algumas vezes, é chamado de 'causação descendente' (*downward causation*). Diversos debates têm ocorrido sobre como a noção de 'causação descendente' deveria ser interpretada, de modo a fazer algum sentido (e.g. Kim 1993). Concordamos com El-Hani & Emmeche (2000: 262), que afirmam que 'causação descendente' pode ser entendida como uma expressão aristotélica de causalidade formal: 'Entidades de níveis superiores estabelecem um padrão particular de restrições sobre as relações de entidades de níveis inferiores que os compõem.'

Forças de causação descendentes de padrões de alto nível podem ser observados como restringindo as possibilidades de interação entre componentes em níveis inferiores. Finalmente, o fenômeno de 'difusão causal' pode ser observado em relação aos robôs. O fenômeno é assim definido por Wheeler e Clark (1999: 106): 'O fenômeno de interesse torna-se dependente, de maneiras inesperadas, de fatores causais externos ao sistema.'

É bem conhecido o fato de que a IA tradicional ajusta seu foco para o que acontece *dentro* do sistema. O sistema nervoso central, artificial ou biológico, é visto como o fator causal principal do com-

portamento. Mas, de acordo com Wheeler e Clark, as causas de nosso comportamento não devem ser procuradas dentro de nós, uma vez que se encontram dispersas no ambiente.

Ora, com a finalidade de entender o comportamento dos robôs, é necessário tomar em consideração vários e diversos aspectos de seus corpos e de seus ambientes, além do modo como interagem e se auto-organizam.

AUTONOMIA, *Umwelt* E VIDA

Ziemke e Sharkey (2001: 725-726, 730) examinaram a autonomia e o *Umwelt* de robôs como consideravelmente independentes de seus criadores humanos. Eles estão interessados em robôs que evoluem por meio de algoritmos genéticos e que são controlados por redes neurais recorrentes. De acordo com os autores, tais robôs se adaptam a seus ambientes e possuem uma base histórica em suas reações. Isto é, as reações dos robôs são 'subjetivas' porque eles estão se auto-organizando, uma vez que não estão completamente construídos, e porque são específicas, conforme as histórias de suas experiências. Além disso, os robôs estão envolvidos em processos sígnicos, fazendo uso de signos, o que os provê com um certo nível de autonomia epistêmica. Como afirmam Ziemke e Sharkey, robôs 'se viram por conta própria' quando interagem com o ambiente. O desenvolvimento dos controladores de robôs (e.g. redes neurais artificiais) e, algumas vezes, de seus corpos (em casos de simulação) seguem aquilo que von Uexküll chamou de 'princípios centrífugos'. Eles se desenvolvem de dentro para fora, contrariamente aos princípios centrípetos mais comuns, de partes pré-arranjadas em conexão (como o braço de um robô ou um sensor ótico) com uma unidade central, de fora para dentro. Finalmente, robôs podem co-evoluir com outras entidades em desenvolvimento.

Ziemke e Sharkey mencionam, como exemplo, o trabalho de Nolfi e Floreano (1998) em que robôs (*kheperas*), controlados por redes neurais recorrentes, co-evoluem com outros robôs, exibindo comportamento de predador ou presa. Cliff e Miller (1996) fornecem um exemplo de co-evolução interna, em que o controlador e o sensor ótico evoluem de um modo cruzado.

Considerando estes exemplos, podemos inferir que há boas razões para suspeitar que robôs estão bem qualificados para apresentar, em

algum nível, autonomia e *Umwelt*. Isto é, somos capazes de fornecer subsídios para afirmar que, de um modo rudimentar, robôs fazem coisas 'por si mesmos' e precisam ter um mapeamento de ordem superior e uma avaliação de amostras sensório-motoras de seu ambiente. Entretanto, Ziemke e Sharkey (2001: 736) terminam seu artigo com um claro 'não' à indagação sobre a possibilidade de *Umwelt* em robôs, exatamente porque não estão vivos: 'Os componentes poderiam estar melhor integrados, depois de serem auto-organizados. Eles poderiam mesmo ser considerados 'mais autônomos' por esta razão, mas eles certamente não se tornaram vivos neste processo'. O mesmo veredicto é dado por Emmeche (2001: 19):

> O que dá ao *Umwelt* seu caráter fenomenal não é o aspecto cibernético-funcional de processamento de sinais, dentro do sistema (e na interface sistema-ambiente), mas o fato de que o organismo vivo é antes constituído como um sujeito ativo com algum agenciamento. Portanto, podemos dizer que apenas seres vivos genuínos (organismos e animais, especialmente) vivem a experiência de um *Umwelt* .

Então, robôs não possuem *Umwelt* porque não estão vivos e não se tornam vivos em sua interação crescente e autônoma com o mundo. Se este argumento está correto, desqualifica imediatamente criaturas artificiais e exige da Robótica que ela se torne um ramo da biologia na tentativa de produzir criaturas com *Umwelt*.

Neste ponto, entretanto, gostaríamos de levantar uma questão que pode parecer estranha (ao menos assim me pareceu quando pensei a seu respeito pela primeira vez): o que a vida tem a ver com isto? Em primeiro lugar, vida e experiência não são sinônimos. A possibilidade (ou não possibilidade) de haver experiência sem vida é uma questão empírica. Também é uma questão empírica se criaturas artificiais (para serem distinguidas de criaturas vivas) podem ter *Umwelt*. As pesquisas em Robótica estão destinadas a investigar as capacidades e as propriedades dos robôs. A experiência de que um *Umwelt* pode surgir em determinados tipos de criaturas, vivas ou artificiais, poderia ser uma questão de *descoberta*, devido aos efeitos emergentes de acoplamento entre sistemas de controle (e.g. cérebros), corpos e ambientes. Isto não quer dizer que não existam diferenças entre criaturas vivas e artificiais; quer dizer, apenas, que a capacidade de ter experiências pode não constituir a diferença entre elas.

Em segundo lugar, existem situações em que podemos dizer que criaturas vivem sem experiências. O sono profundo sem sonho é

considerado, normalmente, sem experiências, assim como algumas formas de coma. Organismos sem um sistema nervoso também são, geralmente, assim considerados (e.g. Damasio 1999; Emmeche 2001). Portanto, estar vivo não é suficiente para ter experiências.

Mais difícil, obviamente, é conceber o contrário, que criaturas têm experiências embora não estejam vivas. A sugestão de que uma criatura que não está viva possa ter experiências certamente parece bizarra. Há uma forte tendência para associar 'não vivo' com 'morto', e 'estar morto' com o estado de não experimentar coisa alguma. Entretanto, parece que, em relação a robôs, esta associação não é adequada. Basicamente, o que estamos propondo é que criaturas artificiais não se enquadram perfeitamente nas categorias 'morto' ou 'vivo'. Não podemos decidir a respeito de suas capacidades, com base na tentativa de forçá-las a uma dessas classes.

Então, a vida é uma condição necessária para a experiência? Por quê? Quais são exatamente os argumentos? Há, pelo menos, dois argumentos a serem examinados. De acordo com o primeiro argumento, todas as criaturas capazes de experiência que conhecemos são criaturas vivas. Um testemunho da tendência para associar vida com experiência são as experiências relatadas de vida após a morte. Em resposta, só podemos recorrer à refutabilidade indutiva. Pode ser o caso que, até hoje, todas as criaturas em relação às quais podemos dizer que 'têm experiência' são criaturas vivas. Mas isto não constitui uma prova para a tese de que a vida é uma condição necessária para a experiência. A não-confiabilidade do argumento indutivo é ainda mais imperativo porque a Robótica dedica-se à construção de criaturas de um tipo totalmente novo, artificial e não-biológico. Exatamente quais propriedades devem ser atribuídas a tais criaturas não é algo que pode ser decidido com base em experiências passadas.

Um argumento mais substancial é aquele que poderia explicar *porquê criaturas* que tem experiências *devem* estar vivas. Tal argumento destaca que a matéria das criaturas vivas é necessária para a *constituição* de experiências. Propomos um olhar mais atento sobre este assunto. Que propriedades da matéria viva seriam essenciais para caracterizar a experiência?

DA MATÉRIA PARA EXPERIÊNCIA: AUTOPOIESIS

Não encontramos argumento que relacione diretamente a capacidade de ter experiências às propriedades específicas da matéria. Um conceito, entretanto, que tem sido freqüentemente mencionado como sendo essencial para a matéria viva é *autopoiesis* (e.g. Maturana & Varela 1987). Ele se refere à capacidade de autogeração e auto-manutenção de células, por exemplo. A questão que se coloca aqui é: como a autopoiesis pode estar relacionada à experiência?

As células são consideradas importantes exemplos de sistemas autopoiéticos. Entretanto, elas não são geralmente consideradas como capazes de experiênciar algo. Emmeche (2001: 18), por exemplo, afirma que protozoários não possuem um sistema nervoso e não detêm um *Umwelt* genuíno. Pode ser que a autopoiesis não constitua um pressuposto para a experiência. Além disso, a noção de autopoiesis não reflete algumas qualidades intrínsecas de um tipo específico de matéria, mas indica uma característica de sua organização, denotando, portanto, uma qualidade funcional. Como afirmam Maturana e Varela, (1987: 51; ver também Ziemke e Sharkey 2001: 732): '... o fenômeno que eles geram, funcionando como unidades autopoiéticas, depende de sua organização e do modo como esta organização acontece, não da natureza física de seus componentes'.

Conclui-se que outras matérias, além da matéria viva, podem instanciar autopoiesis. O material particular que compõe o organismo vivo pode ser resultado de coincidência histórica ou evolucionária:

> A organização autopoiética pode ser obtida por diferentes tipos de componentes. Temos de perceber, entretanto, que, em consideração à origem molecular de seres vivos terrestres, somente certas espécies moleculares são provavelmente possuidoras das características necessárias para unidades autopoiéticas, desta forma iniciando a história estrutural à qual pertencemos (Maturana e Varela 1987: 49).

Se *Umwelt*, a experiência subjetiva de interação perceptiva e motora com o ambiente, ocorre na dependência de certas características *funcionais*, normalmente encontradas em algumas formas de matéria, porque essas características funcionais não podem ser replicadas em outros materiais? Longe de desqualificar robôs, este argumento nos leva a considerar mais cuidadosamente a relação entre função, forma e matéria das criaturas artificiais.

FUNÇÃO, FORMA E MATÉRIA

No cerne da ciência cognitiva está, tradicionalmente, a idéia de que qualquer sistema deveria ser analisado em três diferentes níveis. No nível funcional, a tarefa do sistema é examinada e hipóteses são formuladas sobre o que o sistema deve fazer. No nível formal, ou computacional, são estudadas as maneiras como a função pode ser desempenhada. Neste nível, a forma dos processos está sob investigação. Finalmente, no nível físico, considera-se a composição material do sistema que está sendo analisado. Iremos usar o termo 'função', 'forma' e 'matéria' para fazer referência aos três níveis de análise.

Da perspectiva tradicional da IA, tendo em vista os três níveis de análise, a situação é aproximadamente esta. No nível funcional, considera-se a tarefa que o (sub)sistema tem de cumprir. A análise, neste nível, resulta na formulação do problema que o sistema deve resolver através de processamento interno de informação. Marr (1982), por exemplo, afirma que a tarefa a ser resolvida pelo sistema visual seria a transformação de uma imagem *retiniana* bidimensional em um mundo tridimensional de objetos reconhecidos. No nível da organização, a questão está na relação computacional entre as representações: algoritmos especificam o modo através do qual os problemas podem ser resolvidos. Diferentes algoritmos podem ser comparados a partir de suas performances e/ou plausibilidades psicológicas, etc. No nível físico, para a IA tradicional, é aceitável praticamente qualquer coisa, porque algoritmos podem 'rodar' em um vasto espectro de construções e materiais, de computadores *high techs*, até latas de cerveja, como Searle tem indicado. Poucas limitações são impostas ao nível físico. Qualquer substrato material que possa implementar processos formais é, em princípio, suficiente.

Para a *nouvelle* IA, a situação é diferente. No nível funcional, o sistema deve atingir homeostase, tentando manter sua própria organização enquanto interage com o ambiente. Em relação à forma de organização, e ao modo como seus componentes interagem, o foco da abordagem é direcionado à auto-organização do sistema e àquilo que permite um acoplamento dinâmico entre sistema e ambiente. As condições para homeostase e auto-organização resultam em coerções mais restritivas sobre a matéria na qual estes processos são encontrados, implicando pré-requisitos para as unidades

autopoiéticas. A seguinte tabela (tabela 1) sumariza as diferenças entre a tradicional e a *nouvelle* IA:

	Tradicional IA	*Nouvelle* IA
Função (nível tarefa)	solução de problemas	Homeostase
Forma (nível organizacional)	algoritmos	auto-organização
Matéria (nível físico)	qualquer coisa	unidades autopoiéticas

Tabela 1: IA tradicional versus *Nouvelle* IA

Nossa sugestão é a de que a *nouvelle IA* leva a uma reconsideração das respostas fornecidas pela ciência cognitiva tradicional nos três níveis de análise. Em parte, a Robótica é importante para a ciência cognitiva porque as idéias básicas apresentadas pela *nouvelle* IA, nos três níveis, resultam em restrições nas relações entre eles. Esta sugestão impõe considerações mais cuidadosas quanto ao material de constituição dos robôs. Essencialmente, isto não somente se aplica ao sistema nervoso artificial dos robôs, mas, muito amiúde, ao material de constituição do robô como um todo.

AS CRIATURAS ARTIFICIAIS PROTO-*UMWELT*

Vamos considerar o que poderia vir a ocorrer se robôs fossem desenvolvidos de acordo com estas idéias. Suponhamos um sistema artificial que consista em unidades autopoiéticas. Elas se desenvolvem através de princípios centrífugos, de dentro pra fora. Seus processos acoplados resultam na emergência auto-organizada de padrões de comportamento que direcionam o sistema à homeostase. O quê diríamos sobre sistemas como estes? Talvez não disséssemos que estão vivos, mas teriam *Umwelt*? Emmeche (2001: 21) nos brinda com a seguinte resposta:

> Se tal dispositivo material, como um robô, pudesse ter a flexibilidade orgânica de um animal, permitindo-o instanciar qualquer coisa como uma lei da mente, isto é, como uma tendência para permitir que signos se influenciem mutuamente, de um modo auto-organizado, é difícil ver porque tais dispositivos não deveriam poder realizar signos genuínos (incluindo *qualisignos*).[4] [...] Se aquele sistema construído artificialmente realiza uma ação de qualisigno, e de senti-

mento vivo, ele teria mente no sentido de Peirce. Mas teria um *Umwelt* ? Há alguém em casa experimentando algo?

Parece injusto repetir a questão depois de ter cedido tanto. Quando há 'ação do *qualisigno*' e 'sentimento vivo', parece que há um *Umwelt*. Em algum ponto a 'questão perene' deveria ser detida, caso não signifique mais do que a repetição do problema das 'outras mentes'. Ao mesmo tempo, pode-se, talvez, ser solidário com a relutância em atribuir uma forma de experiência a um simples robô.

Nestas circunstâncias, talvez possamos evitar a repetição da 'questão perene' da IA, considerando um estágio intermediário, na rota para um *Umwelt* completo. Talvez seja útil qualificar robôs como um tipo de organismo rudimentar; algo que pudesse ter um 'proto-*Umwelt*'. É interessante lembrar que os biólogos, algumas vezes, falam sobre células de maneira similar. Emmeche (2001: 18), por exemplo, ao tratar de protozoários, menciona '...a falta de um sistema nervoso e um *Umwelt* verdadeiro, mas [...] há um 'círculo *autocinético* e móvel', mais simples, pelo qual [protozoários] entram em interações semióticas com seu meio exterior'.

Finalmente, o neurocientista Antônio Damásio (1999: 136-137) sugere que uma ameba, por exemplo, pode ser vista como possuindo 'alguns antecedentes biológicos da noção de *self*'. Assim, há conhecimento de uma forma de autocinese que leva a algo próximo à capacidade semiótica de um *self*, sem criar um *Umwelt*. Uma vez que aceitemos esta possibilidade, torna-se difícil ver porque sistemas artificiais homeostáticos, constituídos por atraentes unidades autopoiéticas em autocinese, e em interação semiótica com seu ambiente, não teriam um 'proto-*Umwelt*' do tipo que é conferido a criaturas unicelulares. Por que não atribuir a criaturas não-vivas o mesmo tipo de 'proto-*Umwelt*', atribuído por biólogos às células?

AGRADECIMENTOS

Gostaria de agradecer à Raquel Pellizzetti, Mariana Broens, João Queiroz e Eunice Gonzalez pela ajuda no desenvolvimento deste texto; à Fapesp pelo apoio financeiro, à UNESP (Campus de Marília), por me oferecer condições de trabalho e ao NICI (Nijmegen), por permitir o desenvolvimento de minhas atividades em São Paulo.

NOTAS

[1] O rótulo ‘IA’, neste contexto, inclui o que Haugeland chama de ‘a Boa e Velha Inteligência Artificial’ ou ‘a Inteligência Artificial à Moda

Antiga' (*Good Old Fashioned Artificial Intelligence — GOFAI*: *the symbol rule approach*) e formas tradicionais de conexionismo, por exemplo, as redes *feedforward*, populares nos anos 80. Esta forma de IA deve ser diferenciada de trabalhos mais recentes em Robótica que, muitas vezes, envolvem pesquisas com agentes autônomos. Entretanto, uma vez que tem sido questionada a autonomia e independência destes agentes, faremos uso da expressão '*nouvelle* IA', como rótulo geral para os trabalhos mais recentes em Robótica.

[2] Estamos deixando de lado aqui questões como 'de que modo iremos saber se robôs têm experiências ou não?' porque a mesma questão poderia ser levantada em relação a experiências de outra pessoa (o conhecido problema 'outras mentes').

[3] Notas dos orgs.: O leitor deve consultar, neste volume, o capítulo de El-Hani & Queiroz, onde é desenvolvido um tratamento cuidadoso sobre a noção de emergência.

[4] Nota dos orgs.: Para Peirce (CP 8.334), o signo, com respeito a sua própria constituição, pode ser uma qualidade (qualisigno), um existente (sinsigno), ou uma lei (legisigno). Um qualisigno é uma qualidade que é um signo, que funciona como um signo sem qualquer referência a qualquer outra 'coisa'.

Referências

EMMECHE, C. 2001. Does a robot have an *Umwelt*? Reflections on the qualitative biosemiotics of Jakob von Uexküll. *Semiotica* 134 (1): 653-693.

BRAITENBERG, V. 1984. *Vehicles:* Experiments in synthetic psychology. MIT Press.

CLARK, A. 2001. *Mindware:* an introduction to the philosophy of cognitive science. Oxford University Press.

CLIFF, D.; MILLER, G. 1996. Co-evolution of pursuit and evasion II: Simulation methods and results. Em: P. Maes, M. Mataric, J.A. Meyer, J. Pollack & S. Wilson (eds). *From animals to animats* 4: Proceedings of the fourth international conference on simulation of adaptive behavior (SAB96). Cambridge, MA: MIT Press. pp. 506-515.

DAMASIO, A. 1999. *The feeling of what happens:* Body, emotion and the making of consciousness. Vintage.

DRAAISMA, D. 1986. De machine achter het gordijn: Drie eeuwen debat over bewustzijn en mechanicisme. Em: P. Hagoort & R. Maessen (eds.). *Geest, computer, kunst*. Grafiet.

EL-HANI, C.; EMMECHE, C. 2000. On some theoretical grounds for an organism-centered biology: Property emergence, supervenience, and downward causation. *Theory of Bioscience* 119: 234-275.

FLEW, A. 1964. *Mind, body, and death*. MacMillan Publishing Company.

KAUFMANN, S. 2000. *Investigations*. Oxford University Press.

KENNY, A. 1970. *Descartes:* Philosophical letters. Basil Blackwell.

KIM, J. 1993. *Supervenience and mind:* selected philosphical essays. Cambridge University Press.

MARR, D. 1982. *Vision*. Freeman.

MATURANA, H.R.; VARELA, F.J. 1987. *The tree of knowledge:* The biological roots of human understanding. Shambhala.

NOLFI, S.; FLOREANO, D. 1998. Co-evolving predator and prey robots: Do 'arm races' arise in artificial evolution? *Artificial Life* 4: 311-335.

SIMMEN, R. 1968. *Mens & Machine:* Teksten en documenten over automaten, androïden en robots. Amsterdam: van Lindonk.

TURING, A.M. 1950. Computing machinery and intelligence. *Mind* 59: 433-460.

UEXKÜLL, J. von. 1936/2001. An introduction to *Umwelt*. *Semiotica* 134: 107-110.

UEXKÜLL, J. von. 1937/2001. The new concept of *Umwelt*: A link between science and the humanities. *Semiotica* 134: 111-123.

WHEELER, M.; CLARK, A. 1999. Genic representation: Reconciling content and causal complexity. *British Journal for the Philosophy of Science* 50: 103-135.

WIT, H.C.D. de 1982. *Ontwikkelingsgeschiedenis van de biologie*, Deel 1. Pudoc.

ZIEMKE, T. e Sharkey, N. 2001. A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica* 134: 701-746.

COMPUTAÇÃO, COGNIÇÃO, SEMIOSE.

JOÃO QUEIROZ
ÂNGELO LOULA
RICARDO GUDWIN
ORGANIZADORES

EDUFBA



CAPÍTULO 10

ENGENHARIA IMUNOLÓGICA E COGNIÇÃO: DA NATUREZA À SOLUÇÃO DE PROBLEMAS DE ENGENHARIA

Leandro Nunes de Castro,Janaína Stella de Sousa, George Barreto Bezerra

INTRODUÇÃO

Computação natural é o termo usado para descrever sistemas computacionais desenvolvidos com inspiração em fenômenos naturais, ou sistemas que usam a natureza como meio para realizar computação (de Castro 2006). Por exemplo, redes neurais artificiais, algoritmos evolutivos, algoritmos de inteligência coletiva (*swarm intelligence*) e, mais recentemente, sistemas imunológicos artificiais são todos *sistemas inspirados na biologia*, um dos ramos da computação natural. Outro ramo é a *biologia motivada pela computação*, composta principalmente pela vida artificial e geometria computacional. Finalmente, *biocomputação*, como, por exemplo, computação molecular, constitui o terceiro ramo da computação natural.

O que todos os ramos da computação natural têm em comum é uma íntima relação com a natureza, levando a pesquisas altamente interdisciplinares. Isso pode trazer resultados bastante promissores para os campos envolvidos, mas traz também algumas dificuldades, pelo menos nos estágios iniciais da pesquisa. Este capítulo introduz a *engenharia imunológica*, ou seja, tem seu foco no sistema imunológico biológico como fonte de inspiração para o desenvolvimento de ferramentas computacionais para a solução de problemas complexos de engenharia. Embora este capítulo enfoque a engenharia imunológica, muitas das idéias apresentadas podem ser vistas como técnicas genéricas de engenharia de algoritmos de computação natural.

ENGENHARIA IMUNOLÓGICA

Os sistemas imunológicos artificiais (SIA) compõem uma nova abordagem da inteligência computacional inspirada nas teorias da imunologia, seus princípios e modelos, com aplicações na solução de problemas. Como toda nova abordagem (e.g. 'inteligência coletiva'; Bonabeau et al. 1999, Kennedy et al. 2001), os sistemas imunológicos artificiais ainda necessitam de uma descrição mais formal e de melhor fundamentação teórica. Entretanto, novas perspectivas já foram apresentadas em um livro recentemente publicado (de Castro & Timmis 2002). Este livro, *Artificial Immune Systems: A New Computational Intelligence Approach*, cobre variados tópicos e domínios, da biologia à computação. Há capítulos dedicados a imunologia, neurociência, e endocrinologia, sempre com uma visão sobre como tais sistemas são importantes para o desenvolvimento de ferramentas computacionais visando a solução de problemas complexos. O livro também oferece uma pesquisa da literatura em SIA e em sistemas híbridos dos SIA com outras técnicas, tais como redes neurais artificiais, sistemas nebulosos, algoritmos evolutivos e outros. O capítulo 2 do livro faz uma revisão do sistema imunológico biológico dos vertebrados e o capítulo 3 introduz o processo de *engenharia imunológica*. O termo *engenharia imunológica* foi criado por Leandro N. de Castro e Fernando J. von Zuben, tendo sido formalizado em uma tese de doutorado (de Castro 2001). Ele se refere a

> um *processo de meta-síntese*, que vai definir a ferramenta de solução de um determinado problema baseado nas caracte-

> rísticas do próprio problema, e depois vai aplicá-la na obtenção da solução. Ao invés de buscar a *reconstrução parcial ou total do sistema imunológico tão fielmente quanto possível*, a engenharia imunológica deve procurar *desenvolver e implementar modelos pragmáticos inspirados no sistema imunológico* que preservem algumas de suas propriedades essenciais e que se mostrem passíveis de implementação computacional e eficazes no desenvolvimento de ferramentas de engenharia (de Castro 2001: 44).

Alguns termos foram destacados (itálico) porque propõem importantes conceitos e idéias. Primeiro, o conceito de meta-síntese, que se refere ao processo no qual um sistema tem a capacidade de adaptar-se em busca da solução de um problema baseando-se em suas interações com o próprio problema (ambiente). Este conceito foi inicialmente proposto no sentido de usar as propriedades do sistema imunológico para desenvolver uma nova técnica construtiva, que incluísse 'etapas de poda', para o projeto de redes neural artificiais.

O sistema imunológico tem grande potencial para gerar um repertório de células e moléculas capazes de combater elementos invasores causadores de doenças, conhecidos como patógenos (e.g. vírus, bactérias e fungos). Através da modificação da estrutura dos receptores das células imunológicas, e do aumento da concentração de determinadas células e moléculas no sangue e na linfa, o sistema imunológico pode também se tornar cada vez mais eficiente em reconhecer e destruir patógenos. Esse sistema é então inerentemente capaz de definir sua própria arquitetura e ajustar seus 'parâmetros' de tal forma a lidar com os invasores. Historicamente, o que se esperava quando se propôs a engenharia imunológica era que, através do estudo de como o sistema imunológico combatia os patógenos, poderíamos usar princípios e processos para projetar novos tipos de algoritmos de aprendizagem para redes neurais artificiais.

Outra questão importante levantada pela definição da engenharia imunológica refere-se a criação de modelos precisos. Sob uma perspectiva de engenharia, é importante considerar quão preciso um modelo precisa ser em relação à sua utilidade como técnica para solucionar problemas. A aplicação da análise matemática e da modelagem pode resultar em alguns benefícios como uma descrição quantitativa de como o sistema imunológico funciona, uma análise mais crítica das hipóteses, pode ajudar na predição de comportamentos, no desenvolvimento de experimentos, interpretação de fenômenos, e assim por diante. Note que o objetivo desses modelos é consideravelmente diferente daqueles para os quais se proje-

tam ferramentas de engenharia para solução de problemas. Como sugerido na definição de engenharia imunológica, nossa preocupação é manter um modelo pragmático e útil como ferramenta de engenharia. Logo, devemos ser cuidadosos para não estender demais a metáfora.

Seleção clonal, expansão e maturação de afinidade

No sistema imunológico dos vertebrados, os processos de seleção e expansão clonal juntamente com a maturação de afinidade, são exemplos claros de um processo meta-sintético. Quando um patógeno invade o corpo, algumas das células imunológicas reconhecem esse patógeno e começam a se proliferar. Como todo processo reprodutivo, a reprodução celular (clonagem) no sistema imunológico está sujeita a erros, denominados 'mutação'. Essa reprodução é um processo mitótico de divisão celular que pode resultar em erros para as células descendentes geradas. Além disso, a taxa de mutação é inversamente proporcional à afinidade que os receptores imunológicos têm com o patógeno reconhecido.

Em resumo, seleção e expansão clonal, juntamente com a maturação de afinidade, ocorrem da seguinte maneira. Nosso sistema imunológico é composto de um grande número de células que apresentam receptores em sua superfície. Estes receptores são responsáveis por se ligar com porções de patógenos, conhecidas como antígenos, e sinalizar para outras células que irão eliminar os patógenos marcados (reconhecidos). Entretanto, os invasores patológicos se replicam em nosso organismo, aumentando cada vez mais o dano causado ao organismo hospedeiro. O sistema imunológico desenvolveu uma maneira para combater a infecção através de uma replicação das células imunológicas de forma a compensar a proliferação dos patógenos. Como discutido, essa replicação de células imunológicas não é perfeita; erros ocorrem a uma taxa inversamente proporcional à qualidade do reconhecimento entre receptores celulares e patógenos reconhecidos. Aquelas células mutadas, cujos receptores possuem alta afinidade com o patógeno, são então selecionadas e mantidas em um repertório chamado 'memória'. Já aquelas que, quando mutadas, não mais reconhecem o patógeno, ou o fazem com uma afinidade muito baixa, possuem grandes probabilidades de serem eliminadas, sofrendo morte ou anergia. A Figura 1 mostra um esquema que resume os processos de expansão clonal e maturação de afinidade.

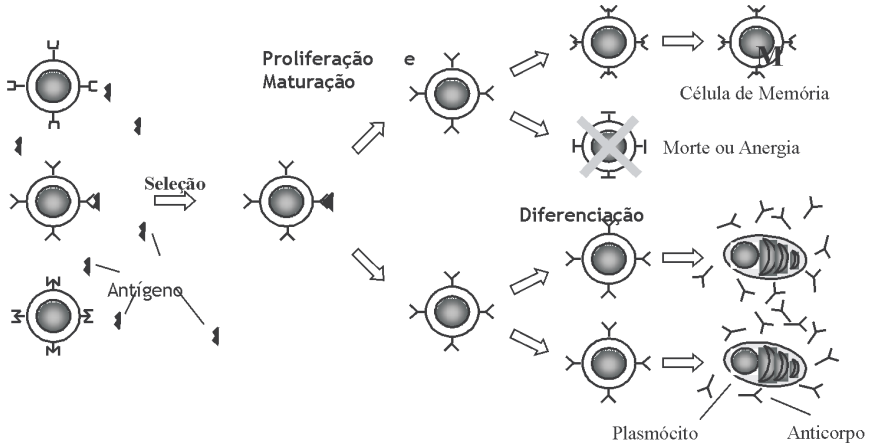


Figura 1: Seleção clonal, expansão e maturação de afinidade. As células B têm receptores em suas superfícies os quais permitem que elas reconheçam os antígenos. Depois do reconhecimento, algumas células são estimuladas para reproduzir; um processo sujeito a erro (mutação). Algumas células que melhoram seus receptores de reconhecimento são selecionadas para se tornarem células de memória, isto é, células com vida prolongada. Outras células que pioram consideravelmente o reconhecimento patogênico tendem a ser eliminadas.

Certamente essas não são as únicas características da seleção clonal interessantes para o processamento de informação, mas elas servem para a proposta original da engenharia imunológica. Por exemplo, sabe-se que a resposta imunológica (seleção clonal) é local, isto é, a resposta não envolve todo o repertório de um determinado tipo de célula imunológica; apenas uma amostra das células está envolvida. Outros pesquisadores desenvolveram diferentes modelos da seleção clonal usando essas idéias (Forrest et al. 1993) com aplicações em outros contextos tais como busca multi-modal.

Como proposto na definição de engenharia imunológica, a idéia é 'desenvolver e implementar modelos pragmáticos inspirados no sistema imunológico'. Sendo assim, os mecanismos básicos da seleção clonal discutidos acima foram suficientes 'para o desenvolvimento ferramentas de engenharia'.

Fazendo engenharia com o princípio da seleção clonal

É interessante perceber que a idéia de meta-síntese está completamente incorporada nos processos de seleção clonal e maturação de afinidade. Não só a estrutura do repertório imunológico vai sendo ajustada ao 'problema' com o qual o sistema imunológico se depara (patógenos), mas também a estrutura das moléculas vão

sofrendo ajustes finos através de mutações somáticas seguidas de seleção. Esses processos, então, servem ao propósito de usar idéias para desenvolver algoritmos de projeto automático e aprendizagem para redes neurais artificiais. A questão que ainda persistia era como usá-las em um contexto de redes neurais.

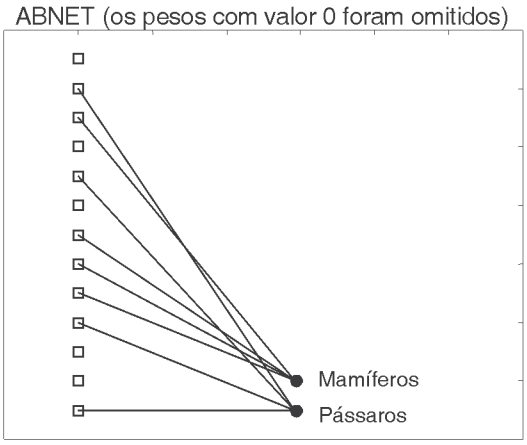
Em um extenso artigo, algumas sugestões de como usar idéias do sistema imunológico para desenvolver novas estruturas de redes neurais artificiais e algoritmos de aprendizagem de máquina foram apresentadas (de Castro et al. 2003). Uma rede neural Booleana construtiva foi proposta baseada nessas idéias. Dado um conjunto de amostras de entrada, e um conjunto de neurônios artificiais Booleanos, isto é, neurônios com vetores de pesos binários, o problema era como determinar uma arquitetura de rede neural apropriada e seu respectivo conjunto de pesos associados, baseado no problema, de forma que ele fosse solucionado satisfatoriamente. A idéia é relativamente simples se tivermos a seleção clonal e a maturação de afinidade em mente.

Começando com uma rede composta por um pequeno número de neurônios, selecione aquele com maior afinidade para um dado padrão de entrada. Os padrões de entrada são apresentados seqüencialmente à rede. O neurônio selecionado é então clonado (reproduzido com uma pequena mutação), e o neurônio descendente com maior afinidade ao padrão de entrada é selecionado para substituir aquele que o gerou, ou para ser adicionado à rede. A afinidade neste caso corresponde à menor distância ao padrão de entrada e também à alta concentração de antígenos, isto é, padrões de entrada. A idéia é então reproduzir aqueles neurônios da rede capazes de reconhecer um grande número de padrões e apresentar a menor distância a esses padrões. Os neurônios da rede que não são reforçados por nenhum antígeno tendem a ser removidos da rede, simulando assim a morte de células não estimuladas no sistema imunológico.

Embora essas idéias possam ser usadas para qualquer tipo de rede auto-organizável, a implementação original foi restrita a redes Booleanas. Dado o conjunto de dados 'Animals' na Figura 2(a), uma rede, chamada ABNET (*AntiBody NETwork*), gerada pelo algoritmo descrito acima é mostrada na Figura 2(b). Note que a rede final é capaz de agrupar as duas maiores classes de mamíferos e pássaros contidos no conjunto de dados sem utilizar nenhuma informação sobre como o conjunto de dados está organizado.

		Dove	Hen	Duck	Goose	Owl	Hawk	Eagle	Fox	Dog	Wolf	Cat	Tiger	Lion	Horse	Zebra	Cow
É	Pequeno	1	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0
	Médio	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
	Grande	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
Possui	Duas pernas	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
	Quatro pernas	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
	Pêlos	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
	Cascos	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
	Crina/Juba	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0
	Penas	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Gosta de	Caçar	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0
	Correr	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	0
	Voar	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
	Nadar	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0

(a)



(b)

Figura 2: A ABNET (rede de anticorpo) quando aplicada ao conjunto de dados 'Animals'. (a) Conjunto de dados de animais. (b) Uma das cadeias geradas pelo algoritmo inspirado na teoria da seleção clonal de respostas de anticorpo.

O princípio da seleção clonal não é apenas útil para projetar redes neurais, o que já poderia ser considerado uma importante contribuição para as pesquisas em inteligência computacional. Qualquer leitor familiarizado com a teoria da evolução e/ou algoritmos evolutivos poderia achar a seleção clonal muito parecida com um

processo evolutivo. De fato a expansão clonal, seleção e maturação de afinidade são semelhantes a um processo *micro-evolutivo*. A expansão clonal é um processo de reprodução, e a maturação de afinidade corresponde a uma variação genética seguida de seleção. Portanto, seleção clonal e maturação de afinidade constituem um processo evolutivo que ocorre em uma escala de tempo muito mais rápida do que a evolução das espécies.

Há, entretanto, algumas diferenças entre a macro-evolução (evolução das espécies) e a micro-evolução (evolução dentro do organismo). Macro-evolução nos organismos envolve diferentes seqüências de passos e mecanismos de variação genética quando comparada com a micro-evolução. Por exemplo, não há *crossover* durante a divisão celular, e a taxa de mutação dos organismos não é proporcional ao seu *fítness*, apenas a taxa de reprodução é.

Outro aspecto importante que chama a atenção sobre o princípio da seleção clonal é o fato de que ele permite que o sistema imunológico aprenda os padrões antigênicos apresentados. Na verdade, este princípio está incorporado no processo de vacinação. O organismo é inoculado com amostras de patógenos enfraquecidos ou mortos, de tal forma que é construído um repertório de células e moléculas imunológicas capazes de reconhecer o patógeno já apresentado anteriormente, antes que ele cause danos no organismo. O problema da resposta imunológica, via seleção clonal, é que o processo de proliferação celular demanda algum tempo, denominado de *atraso da resposta*, da ordem de dias, até que um número suficiente de células e moléculas seja produzido para combater a infecção. Isso pode resultar na apresentação de sintomas (e.g. febre, cansaço, irritações da pele, etc., dependendo do tipo de agente infectante) até que o patógeno seja eliminado. Se o sistema imunológico tiver um primeiro contato com os patógenos enfraquecidos ou mortos, então haverá a fase de atraso que será bem pequena, de forma que os sintomas da doença não se manifestem.

Com essa visão de reconhecimento via receptores imunológicos e aprendizagem através da seleção clonal, é possível desenvolver uma outra ferramenta de engenharia para a solução de problemas de reconhecimento de padrões. O algoritmo CLONALG (*Clonal Selection Algorithm*) foi implementado com esse objetivo. Conceitualmente simples, este algoritmo envolve os mecanismos descritos abaixo (de Castro e Von Zuben 2000):

• Gere um conjunto de soluções candidatas (correspondendo ao repertório de células imunológicas e moléculas);

• Determine os *n* melhores indivíduos da população baseado em sua afinidade com o padrão de entrada (correspondente aos antígenos);

• Reproduza (crie cópias de) os *n* melhores indivíduos proporcionalmente à sua afinidade (correspondendo à fase de expansão clonal);

• Mute essas cópias (clones) de acordo com a afinidade; quanto maior a afinidade, menor a taxa de mutação, e vice-versa;

• Re-selecione os indivíduos mutados de acordo com sua afinidade (processos de maturação de afinidade e seleção).

Note que esse algoritmo é de fato um novo tipo de algoritmo evolutivo inspirado no sistema imunológico. Ele engloba os três principais processos evolutivos de reprodução, variação genética e seleção. Há também similaridades com estratégias evolutivas e técnicas de programação genética. Apesar das semelhanças, a seqüência de passos não é a mesma, e o mais interessante é que o desempenho destes algoritmos para a solução de problemas é qualitativamente diferente. O algoritmo de seleção clonal encontra diversas aplicações em problemas de busca multi-modal, combinando exploração com explotação do espaço de buscas. Por outro lado, os algoritmos evolutivos em geral são muito bons para busca exploratória.[1]

O sistema imunológico tem que lidar com vários tipos diferentes de antígeno; ele não pode privilegiar o reconhecimento de nenhum antígeno em detrimento de outros. O sistema imunológico sugere que um algoritmo imunológico, tal como o CLONALG descrito acima, tem que ser capaz de gerar um repertório de células capaz de cobrir a maioria dos picos de uma região de afinidade conceitual, o que poderia corresponder aos antígenos invasores. Este tipo de comportamento é sempre esperado de um algoritmo desenvolvido com inspiração no princípio da seleção clonal e maturação de afinidade. Isso é exatamente o que propõe a engenharia imunológica: explorar a essência do sistema imunológico para o desenvolvimento de algoritmos que sejam capazes de qualitativamente reproduzir os comportamentos observados no sistema biológico natural.

Com algumas modificações, esse algoritmo foi adaptado para realizar buscas multi-modais, e, como esperado, se mostrou muito eficiente em determinar e manter múltiplos ótimos da região de afinidade (*fitness*). Em contraste, algoritmos evolutivos padrões requerem o uso de esquemas de restrição no cruzamento dos indivíduos, compartilhamento de fitness, mecanismos de aglomeração e outros processos para conseguir cobrir os múltiplos ótimos da superfície de fitness. A Figura 3 ilustra o comportamento típico do CLONALG quando aplicado à otimização de funções multi-modais contínuas.

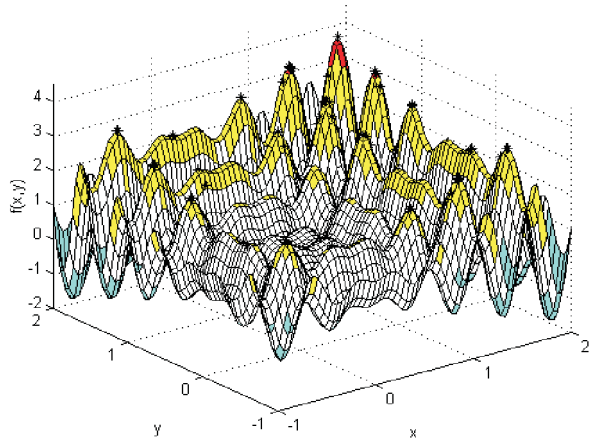


Figura 3: O algoritmo da seleção clonal aplicado à busca multi-modal. Observe a presença de indivíduos (estrelas) em vários picos da superfície.

TEORIA DA REDE IMUNOLÓGICA

Em 1974, Niels Jerne formalizou o que até hoje é conhecido como a teoria da rede imunológica. Sua grande idéia foi perceber que o sistema imunológico não é apenas um sistema reativo que permanece em repouso até que um antígeno invada o organismo. Ele sugeriu que algumas porções dos receptores das nossas células imunológicas poderiam ser reconhecidas por outras células e moléculas do sistema imunológico. Isso poderia resultar em um sistema que é sempre dinâmico, isto é, um sistema que não espera estímulos externos para agir.

Seguindo este esquema, surge então uma importante questão: 'Se o sistema imunológico é capaz de reconhecer suas próprias células e moléculas, porque ele não reage contra o próprio organismo?'

A sugestão proposta até então era que um mecanismo supressivo poderia controlar o reconhecimento das 'células próprias', enquanto um mecanismo de ativação controlaria o reconhecimento do 'não-próprio'. Entretanto, estes mecanismos não foram claramente considerados na teoria e muito menos observados em laboratório, embora algumas referências existam dando suporte biológico para a teoria da rede. A teoria da rede imunológica gerou muito debate em imunologia teórica e experimental. Deve ser destacado, entretanto, que Jerne ganhou um prêmio Nobel em 1984 pela sua teoria e por muitas outras contribuições à imunologia.

Em resumo, a teoria da rede sugere que as células e moléculas imunológicas são capazes de reconhecer a si mesmas e aos antígenos. Este reconhecimento vai resultar em variações nas concentrações e na afinidade (estrutura do DNA) dos receptores imunológicos. Essas variações são funções de vários fatores: 1) os efeitos supressivos da rede, 2) os efeitos de ativação da rede, 3) a morte de células não estimuladas, e 4) o recrutamento de novas células e moléculas para compor o repertório imunológico. A Figura 4 ilustra a teoria da rede imunológica proposta por Jerne (1974).

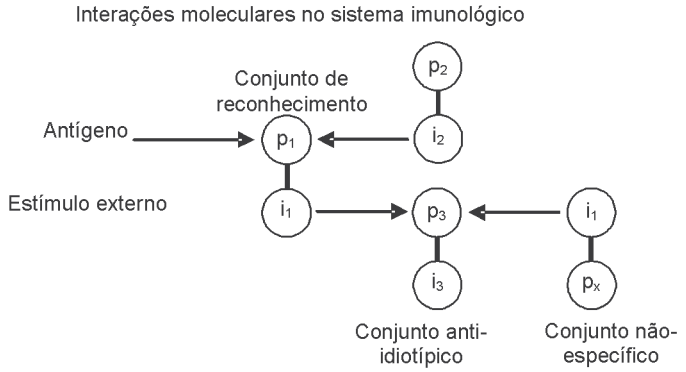


Figura 4: Cada receptor imunológico tem duas porções, p e i. A porção p é capaz de reconhecer outras moléculas, e a porção i pode ser reconhecida por outras moléculas.

Fazendo engenharia com a teoria da rede imunológica

A teoria da rede imunológica parece bastante atraente para qualquer pesquisador em inteligência computacional. Primeiro ela sugere um sistema dinâmico capaz de apresentar interações com ele

próprio e com o meio externo. Segundo, a capacidade de ajustar a estrutura do sistema (rede) e os seus parâmetros às condições adversas do ambiente é uma propriedade muito interessante sob uma perspectiva de engenharia.

Novamente, surge uma outra inspiração do sistema imunológico para o desenvolvimento de uma ferramenta computacional de engenharia. É mais natural olhar o sistema imunológico como uma espécie de dispositivo de reconhecimento de padrões; assim como a primeira versão do algoritmo da seleção clonal discutido anteriormente. O mesmo pode acontecer com a teoria da rede imunológica. A idéia era então implementar uma 'rede imunológica artificial' capaz de realizar reconhecimento de padrões. Isso pode ser natural para um pesquisador com conhecimentos em redes neurais artificiais, pois as RNAs são conhecidas por serem boas em resolver problemas de reconhecimento de padrões e de aproximação de funções.

Imunologistas teóricos já modelaram a rede imunológica empregando equações diferenciais ordinárias para considerar as variações nas concentrações e algumas vezes na afinidade das células imunológicas. A idéia que tínhamos em mente, no entanto, era desenvolver uma rede imunológica mais semelhante a uma rede neural, ou seja, que iria se adaptar aos estímulos de entrada de acordo com um procedimento iterativo. A visão era, portanto, de uma dinâmica discreta e não contínua. Entretanto, a dinâmica da maioria dos modelos de rede imunológica, incluindo a desta nova proposta, contém os seguintes passos básicos:

Taxa de variação populacional	=	Estimulação da rede	-	Supressão da rede	+	Inserção de novos elementos	-	Morte dos elementos pouco estimulados

Um modelo de rede imunológica artificial foi então proposto, incorporando todos os passos descritos na equação acima. O algoritmo pode ser resumido da seguinte forma:

1. *Inicialização:* crie uma população inicial aleatória de células;

2. *Apresentação dos antígenos:* para cada padrão (antigênico) de entrada faça;

 2.1 *Seleção clonal e expansão:* para cada célula da rede, determine sua afinidade com a entrada apresentada. Selecione um número de células de alta afinidade e reproduza-as (clone) proporcionalmente a sua afinidade;

2.2 *Maturação de afinidade:* aplique em cada um dos clones uma mutação inversamente proporcional à afinidade. Re-selecione um número dos clones de maior afinidade e coloque-os num conjunto de clones de memória;

2.3 *Morte dos elementos não estimulados:* elimine todos os clones de memória cuja afinidade com o antígeno é menor que um limiar pré-definido;

2.4 *Interações clonais:* determine as interações da rede (afinidade) de todos os elementos do conjunto de memória clonal;

2.5 *Supressão clonal:* elimine aqueles clones de memória cuja afinidade um com o outro é maior que um limiar pré-definido;

2.6 *Construção da rede:* incorpore os elementos restantes do conjunto de clones de memória juntamente com todas as células da rede;

3. *Interações da rede:* determine a similaridade entre cada par de células da rede;

4. *Supressão da rede:* elimine todas as células cuja afinidade é maior que um limiar pré-definido;

5. *Introdução de novos elementos:* introduza um número de novas células geradas aleatoriamente na rede;

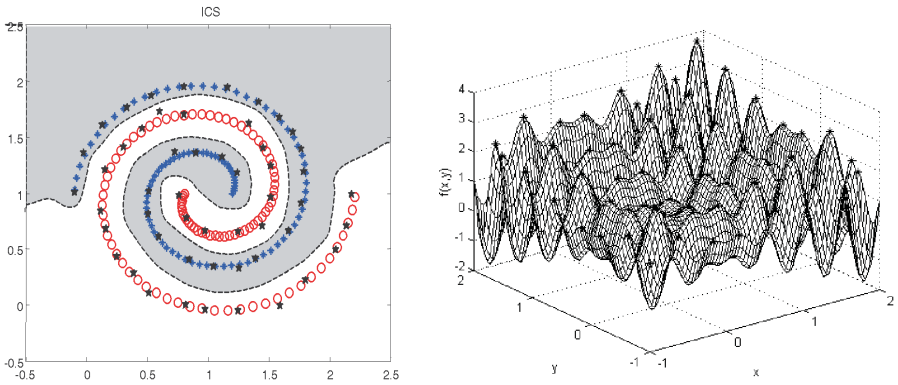
6. *Ciclo:* repita os passos 2 a 5 até atingir um número pré-especificado de iterações.

Algumas características desse algoritmo merecem comentários. Primeiro, note que o algoritmo de seleção clonal e o da maturação de afinidade estão incorporados nos passos 2.2 e 2.3, respectivamente. Além disso, as interações da rede, passos 2.4 a 4, permitem que a rede controle automaticamente o seu número de células.

Esse algoritmo demonstrou ser muito eficiente para realizar compressão de dados. Juntamente com um simples (embora poderoso) procedimento da teoria de grafos, chamado árvore geradora mínima (Zahn, 1971), foi demonstrada também a sua utilidade como técnica de clusterização de dados.[2]

Sem a ajuda da árvore geradora mínima, o algoritmo foi eficientemente aplicado na determinação automática de funções de base radial para redes neurais do tipo RBF (de Castro e Von Zuben 2001b).

Algumas variações do algoritmo padrão também foram aplicadas à otimização de funções multi-modais. A Figura 5(a) e (b) ilustra o desempenho da rede quando utilizada em combinação com uma rede neural do tipo RBF para realizar classificação de padrões e quando aplicada a problemas de otimização, respectivamente.



(a) (b)

Figura 5: Um modelo de rede imunológica artificial aplicada para definir os centros das funções de bases radiais da rede neural (a), e o resultado da otimização de uma função multi-modal (b).

DIVERSIDADE NO SISTEMA IMUNOLÓGICO

Uma questão que pode intrigar muitas pessoas que estudam o sistema imunológico está relacionada a como, com uma quantidade finita de células e moléculas, o sistema imunológico é capaz de detectar um número quase ilimitado de antígenos? Não há apenas uma única resposta para essa pergunta. Primeiro, cada antígeno tem um certo número de diferentes porções da sua superfície que permitem que ele seja reconhecido por mais de um receptor. Segundo, embora todos os receptores de uma determinada célula imunológica tenham a mesma especificidade (isto é, reconhecem um único tipo de antígeno), a diversidade de receptores é extremamente grande no sistema imunológico. Entretanto, essa diversidade intrínseca de receptores traz à tona uma outra questão: como esses receptores são gerados?

Sabe-se que os receptores celulares são gerados através de recombinações aleatórias de segmentos de DNA distribuídas em bibliotecas de genes contidas no próprio DNA das células. Esses

genes são responsáveis por determinar as cadeias de polipeptídeos formadoras da estrutura da região de ligação dos receptores celulares, que se ligam por complementaridade à superfície dos antígenos. Essa recombinação produz uma grande variedade de receptores, agindo na geração das células B, e sendo o estágio inicial de introdução de diversidade no sistema. Juntamente com essa recombinação de seqüências de DNA, a mutação com taxas elevadas que ocorre na fase de expansão celular é um outro fator que também contribui para o aumento da diversidade do sistema imunológico. Ela age alterando bases nucleotídicas individuais e, seguida de um rigoroso mecanismo de seleção, faz um ajuste fino na região de ligação dos receptores, criando moléculas que melhor se encaixem à superfície dos antígenos. A combinação desses dois mecanismos — recombinação e mutação — faz com que o sistema imunológico seja capaz de sintetizar um número quase infinito de receptores celulares, mesmo possuindo um genoma finito.

Muitos pesquisadores têm usado algoritmos evolutivos no estudo de efeitos da evolução sobre a codificação de DNA para a síntese de anticorpos (Hightower et al. 1995; Perelson et al. 1996; Oprea 1999). Queríamos estudar a diversidade em populações de indivíduos para testá-la na fase de inicialização de redes neurais artificiais do tipo *feedforward*. No entanto, ao invés de estudar a diversidade usando cadeias binárias ou bases nucleotídicas (A,C,T,G), o nosso objetivo era utilizar vetores de valores reais para representar as células e moléculas do sistema imunológico. Uma solução para criar diversidade em uma população de vetores de valores reais é usar o algoritmo de *simulated annealing* (Kirkpatrick et al. 1987) e definir uma medida de energia capaz de indicar a diversidade da matriz composta pelos vetores iniciais a serem usados no treinamento das redes neurais tipo *feedforward*. Os resultados apresentados foram encorajadores (de Castro & Von Zuben 2001c). Veja na Figura 6 o desempenho médio do método proposto, denominado INIT, quando comparado a outros métodos da literatura aplicados a vários problemas clássicos e de mundo real.

Essa última aplicação de idéias extraídas do sistema imunológico para desenvolver ferramentas computacionais para a solução de problemas, embora com menos inspiração no sistema imunológico, era ainda motivada pela diversidade das células e moléculas imunológicas.

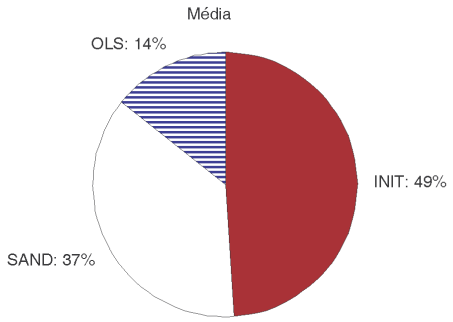


Figura 6: Desempenho médio do algoritmo de geração de diversidade quando comparado com cinco outros algoritmos de rede *feedforward*. Apenas três dos cinco algoritmos apresentaram os melhores resultados médios quando aplicados para 3 problemas *benchmark* e para outros três problemas de mundo real. Em 49% dos casos, o algoritmo proposto, denominado INIT, apresentou um desempenho médio superior aos outros. (ver de Castro & Von Zuben 2001c.)

PROJETO DE SISTEMAS IMUNOLÓGICOS ARTIFICIAIS

Até agora a discussão concentrou-se na extração de idéias de um sistema natural com o objetivo de desenvolver sistemas computacionais para a solução de problemas. O conceito de engenharia imunológica foi discutido e demonstrou ser útil em todos os algoritmos apresentados. Esse processo é ainda novo na comunidade de sistemas imunológicos artificiais, portanto essa terminologia não tem sido amplamente utilizada.

Para concluir o texto, destacaremos pontos importantes para o projeto de um sistema imunológico artificial (SIA), isto é, para fazer engenharia imunológica. Se o leitor já está familiarizado com outro tipo de abordagem de computação inteligente, ou técnica de solução de problemas, não vai encontrar muitas novidades nas diretrizes a seguir, que podem ser muito intuitivas:

Descrição do problema

Identifique todos os elementos que farão parte do sistema imunológico artificial. Isso inclui variáveis, constantes, agentes, funções, e parâmetros necessários para descrever e resolver apropriadamente o problema. Estes não são sempre conhecidos *a priori* e novos componentes podem ser incluídos no sistema em estágios mais avançados do projeto.

Escolha de alguns princípios imunológicos a serem utilizados

Vários princípios imunológicos, modelos e teorias podem ser usados em diferentes contextos para desenvolver ferramentas computacionais para a solução de problemas. Discutimos aqui como o princípio da seleção clonal e a teoria da rede imunológica foram utilizados no desenvolvimento de algoritmos de busca e clusterização.

Projetando o sistema imunológico artificial

Isso envolve alguns aspectos, como decidir quais componentes imunológicos serão utilizados, como representá-los (criar modelos abstratos desses componentes), e a aplicação desses princípios imunológicos (algoritmos) que controlarão o comportamento do sistema.

No algoritmo de seleção clonal e no modelo da rede imunológica descritos aqui, um simples elemento encontrava-se disponível, célula B. Essas células foram representadas através de diferentes cadeias de atributos num espaço de possíveis soluções conhecido como *espaço de formas* (Perelson & Oster 1979). O princípio da seleção clonal e a teoria da rede imunológica foram usados, respectivamente, para reger a forma com que os sistemas (concentração e afinidade dos receptores das células imunológicas) irão variar ao longo do tempo.

Mapeamento reverso do SIA para o problema real

Depois de resolver o problema, algumas vezes é necessário interpretar (decodificar) os resultados apresentados pelo sistema imunológico artificial dentro do domínio original do problema.

O que está por trás desses quatro passos de engenharia imunológica pode ser resumido na seguinte estrutura de engenharia:

- Extração de idéias e metáforas do sistema natural;
- Definição da representação para os componentes do SIA;
- Definição de um conjunto de funções que guiarão as interações dos elementos do SIA com o ambiente e com eles mesmos; e
- Definição de algoritmos imunológicos para reger a dinâmica do SIA.

Discussão

Esse texto apresentou uma visão da engenharia imunológica, um termo concebido por Leandro N. de Castro e Fernando Von Zuben. Foi apresentada uma discussão sobre o que é engenharia imunológica e como ela pode ser utilizada para projetar sistemas imunológicos artificiais. Alguns resultados das ferramentas aqui apresentadas foram incluídos apenas para ilustração.

Espera-se que o leitor tenha notado como pode-se extrair idéias e princípios de um sistema natural, com o objetivo de desenvolver ferramentas de engenharia para solucionar problemas. Um importante conceito discutido foi o de meta-síntese, processo no qual um sistema é capaz de adaptar-se em busca da solução de um problema baseado em suas próprias interações com o problema e nas interações dos elementos do próprio sistema. Não só os parâmetros do sistema são ajustados, mas também sua arquitetura. Há um mecanismo duplamente plástico incorporado na idéia de meta-síntese.

Agradecimentos

Os autores agradecem ao CNPq e a Fapesp pelo apoio financeiro.

Referências

BONABEAU, E.; DORIGO, M.; THÉRAULAZ, G. 1999. *Swarm Intelligence from Natural to Artificial Systems*. Oxford Unviersity Press.

DE CASTRO, L. N. 2006. *Fundamentals of Natural Computing:* Basic Concepts, Algorithms, and Applications. CRC Press LLC.

DE CASTRO, L. N. & Timmis, J. I. 2002. *Artificial Immune Systems:* A New Computational Intelligence Approach. Springer-Verlag.

DE CASTRO, L. N.; VON ZUBEN, F. J. 2000. The Clonal Selection Algorithm with Engineering Applications, *Proc. of GECCO'00* (Workshop Proceedings), pp. 36-37.

DE CASTRO, L. N.; VON ZUBEN, F. J. 2001a. aiNet: An Artificial Immune Network for Data Analysis. Em: *Data Mining:* A Heuristic

Approach, H. A. Abbas, R. A. Sarker e C. S. Newton (eds.). Idea Group Publishing, Capítulo XII, pp. 231-259.

__. 2001b. Automatic Determination of Radial Basis Function: An Immunity-Based Approach. *International Journal of Neural Systems*, Special Issue on Non-Gradient Learning Techniques 11(6): 523-535.

__. 2001c. An Immunological Approach to Initialize Feedforward Neural Network Weights, *Proc. of ICANNGA'01*, pp. 126-129.

DE CASTRO, L. N.; VON ZUBEN, F. J.; de DEUS Jr., G. A. 2003. (prelo) The Construction of a Boolean Competitive Neural Network Using Ideas From Immunology.

DE CASTRO, L. N. 2001. *Engenharia Imunológica:* Desenvolvimento de Ferramentas Computacionais Inspiradas em Sistemas Imunológicos Artificiais, Tese de Doutorado, DCA - FEEC/UNICAMP, Campinas/SP, Brasil, 286 p.

FORREST, S.; JAVORNIK, B.; SMITH, R. E.; PERELSON, A. S. 1993. Using Genetic Algorithms to Explore Pattern Recognition in the Immune System. *Evolutionary Computation* 1 (3): 191-211.

HIGHTOWER, R. R.; FORREST, S. A; PERELSON, A. S. 1995. The Evolution of Emergent Organization in Immune System Gene Libraries, *Proc. of the 6th Int. Conference on Genetic Algorithms*, L. J. Eshelman (ed.). Morgan Kaufmann, pp. 344-350.

KENNEDY, J.; EBERHART, R.; SHI. Y. 2001. *Swarm Intelligence*. Morgan Kaufmann Publishers.

KIRKPATRICK, S.; GELATT Jr., C. D. e VECCHI, M. P. 1987. Optimization by Simulated Annealing. *Science* 220 (4598): 671-680.

JERNE, N. K. 1974. Towards a Network Theory of the Immune *System. Ann. Immunol.* (Inst. Pasteur) 125C, pp. 373-389.

OPREA, M. 1999. *Antibody Repertoires and Pathogen Recognition:* The Role of Germline Diversity and Somatic Hypermutation, Ph.D. Dissertation, University of New Mexico, Albuquerque, New Mexico, EUA.

PERELSON, A. S.; HIGHTOWER, R.; FORREST, S. 1996. Evolution and Somatic Learning in V-Region Genes'. *Research in Immunology* 147: 202-208.

PERELSON, A. S.; OSTER, G. F. 1979. Theoretical Studies of Clonal Selection: Minimal Antibody Repertoire Size and Reliability of Self-Nonself Discrimination. *J. theor. Biol*. 81: 645-670.

ZAHN, C. T. 1971. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers C-20: 68-86*.